Université Blaise Pascal
N° D.U. : 2574

Université d'Auvergne
Années 2012-2015

Ecole Doctorale Sciences de la Vie, Santé, Agronomie, Environnement

N° d'ordre : 667

THESE DE DOCTORAT

Présentée à l'Université Blaise Pascal pour l'obtention du grade de

DOCTEUR D'UNIVERSITE

Spécialité : Physiologie et génétique moléculaires

NITROGEN USE EFFICIENCY IN BREAD WHEAT (*T. AESTIVUM* L.): BREEDING & GENE DISCOVERY

Soutenue le 27 mai 2015 par

Fabien CORMIER

Directeur de thèse : Sébastien PRAUD

Co-encadrant de thèse : Jacques LE GOUIS

Composition du jury

Mathilde CAUSSE	Rapporteur
Philippe BRABANT	Rapporteur
Brigitte MANGIN	Examinateur
Jean-François RAMI	Examinateur
Christophe TATOUT	Président
Jacques LE GOUIS	Co-encadrant
Sébastien PRAUD	Directeur de thèse

Travaux réalisés dans le cadre d'une convention CIFRE, avec l'aide de l'ANRT

BIOGEMMA
Centre de recherche de Chappes, Route d'Ennezat, Chappes, France

UMR 1095 INRA / UBP GDEC
5 chemin de Beaulieu, Clermont-Ferrand, France
Université Blaise Pascal, Aubière, France

This PhD thesis (129/2012) was funded by the **ANRT** (Association Nationale de la Recherche et de la Technologie) and **Biogemma**. Phenotypic data were obtained thanks to the support of the **ANR** (Agence National de la Recherche) from the ProtNblé project (06 GPLA016). **INRA** (Institut National de la Recherche Agronomique) participated to travel expenses. **CIMMYT** (International Maize and Wheat Improvement Center) provided visiting student facilities.

PUBLICATIONS

Cormier F, Lafarge S, Praud S, Le Gouis J. A genome-wide identification of epistatic interactions to reconstruct nitrogen use efficiency pathways in wheat (*Triticum aestivum* L.). *In prep*

Cormier F, Foulkes J, Hirel B, Moenne-Loccoz Y, Gouache D, Allard V, Le Gouis J. Breeding for increased Nitrogen Use Efficiency : a review for wheat. Plant Breeding. *Expected submission: April 2015*

Cormier F, Ly D, Praud S, Le Gouis J, Crossa J. Improving genomic prediction using a GWAS-based method to pre-select marker in multi-environment data. *Expected submission: April 2015*

Ly D, Huet S, Gauffreteau A, **Cormier F**, Rincent R, Jannink JL, Charmet G (2015) Genomic selection for adaptation breeding: Whole-genome prediction of the response to environmental stress in bread wheat. *Submitted to Genetics Selection Evolution*

Cormier F, Throude M, Ravel C, Lafarge S, Praud S, Le Gouis J (2015) Detection of natural variants of NAM-A1 in bread wheat. *Submitted to Agronomy*

Cormier F, Le Gouis J, Dubreuil P, Lafarge S, Praud S (2014) A genome-wide identification of chromosomal regions determining nitrogen use efficiency components in wheat (*Triticum aestivum* L.). Theor Appl Genet 127:2679-2693

Cormier F, Faure S, Dubreuil P, Heumez E, Beauchêne K, et al. (2013) A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (*Triticum eastivum* L.). Theor Appl Genet 126:3035–3048

PRESENTATION

*speaker **Cormier F***, Le Gouis J, Lafarge S, Praud S (2015) Defining QTL boundaries from GWAS results. Plant and Animal Genome - session: QTL cloning. San Diego, USA, January 11

Le Gouis J*, **Cormier F**, Faure S, Dubreuil P, Heumez E, et al. (2014) Nitrogen use efficiency in bread wheat: breeding tools and recent genetic progress. Wheat Breeding 2014: Tools, targets and progress. Rothamsted Research, Harpenden, UK, January 30

Cormier F*, Le Gouis J, Lafarge S, Praud S (2014) Nitrogen Use Efficiency: From Field Trial to Gene Discovery. Plant and Animal Genome - session: Engineering NUE. San Diego, USA, January 11

CASTING

DIRECTORS

Sebastien Praud	Jeune Maître Yoda
Jacques Le Gouis	Dalaï-lama hyper-réactif

ACTRESSES / ACTORS

Madame Hebert	Anti-oxydant mentolés
Mickael Throude	Valery Damidot no limit
Agathe Mini	Artiste performeuse
Gaetan Touzy	Chasseur ~~alpin~~ cantalien
Stéphane Lafarge	Flamant rose discovery
Pierre Dubreuil	Coach sportif
Lénaïck Jehannin	Maman
Océane De Dios	Voyagiste
Hien Le	Orthophoniste
Delphine Ly	Experte Bui-buiteuse
José Crossa	Papi Crossa
Nadine Duranton	Reine des chambres froides
Ariane Colin	Break danseuse
Magalie Leveugle	Metaleuse blonde
Groupe méthodo	Fabricants de casse-têtes
Membres des comités de thèse	Aiguilleurs du ciel
Adeptes du 10h30	Amateurs de bon café
Adeptes du 16h30	Politologues sexologues
Adeptes de l'abri-bus	Cheminots
Gens du lino	Chefs de cuisine moléculaire
Gens de la moquette	Torréfacteurs moulineurs
CE biogemma	Centre de loisirs
Convention collective	Robert Badinter
Membres des apéros de la Victoire	Expert en architecture gothique
Membres des barbecs en hauteur	Bronzés mono-face
6.1.Y	Danseurs contemporains
Agros M	Décontractants
Noyau dur	Qui dure
Colocdelille63 1.0	Infirmiers
Colocdelille63 2.0	Cuisiniers psychanalistes
Cormier Moisan	Petite sœur et al.
Ma mère, mon père	Sans qui je ne serais pas là

SUMMARY

FIGURES

TABLES

GENERAL INTRODUCTION

CONTEXT OF THE STUDY

Our work was motivated by a global will to increase knowledge on nitrogen (N) use efficiency (NUE) in wheat as N is the most used fertiliser (*e.g.* in Europe; Fertilizer Europe 2012) and wheat a major crop (FAO 2012). Indeed, since the Green Revolution, yields have increased simultaneously with fertiliser application. Due to environmental damages (Goulding 2004; Pathak et al. 2011) and the link between the cost of energy and the cost of N fertiliser (Rothstein 2007), it clearly appears that this agronomic model is not sustainable. However, demand for grain is still increasing (FAO 2011). Thus, we need to increase the production per area and per quantity of N applied, making research on NUE essential. In this sense, internationnal policies set fertiliser reduction as a priority, implying for example in France, discussions about new modalities of wheat varieties registration. In fact, varieties would be tested at both high N and low N, and a maintained yield at low N could give a bonus in the registration score. Thus, NUE could also become a major breeding issue and seed companies would have to adapt their breeding strategies. Biogemma is a private biotechnology company funded by French seed companies (mainly Limagrain) and a technical agricultural institute (Arvalis). Biogemma is therefore deeply concerned. Regarding French farmers, N fertilisers are the second main expenditure (14 %) in the total cost of production behind amortisation (17 % included in mechanization, Fig. 1). Moreover, this cost of production is very close to the market price (around 200€ kg ha^{-1}; Arvalis institut du végétal). Thus, wheat production is mainly profitable due to agricultural subsidies. In a context of decreasing subsidies, the reduction of N supply could be a solution to increase and/or maintain farmers' benefits.

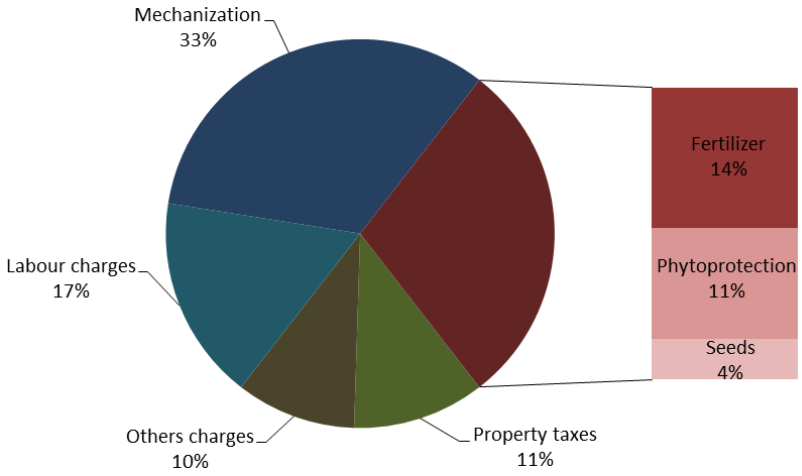


Figure 1: Cost of production in wheat. Arvalis institut du végétal-Unigrain, Cerfrance, 2011.

In 2012, this PhD thesis also started in the context of an important increase in tools and resources dedicated to the research in wheat genetics while the bread wheat genome remains unsequenced due to is complexity (hexaploid, 7 × 3 chromosomes and a size of 17 Gb with 85 % repeated elements). However, from February 2012 to July 2013, the still on-going International Wheat Genome Sequencing Consortium (IWGSC) released high quality genomic sequences for all chromosomes (Eversole et al. 2014), and the

largest chromosome (3B) was the first one for which a reference sequence was produced in 2014 (Choulet et al. 2014). Moreover, using the methodology of Mayer et al. (2011), Biogemma developed a wheat genome zipper which mimics the wheat genome sequence. Its first version was released internally just before the beginning of our work. At the same time, as high-throughput genotyping methods became accessible for most of the members of the wheat community, the amount of available SNP was drastically increased creating the hope that QTL detection, fine mapping and gene cloning would be more easily done in wheat. In this sense, a 90K Illumina SNP chip became available in 2012 (Wang et al. 2014).

All these newly available resources permitted a fresh look on the phenotypic dataset which arose from the ProtNBlé project (2006-2009). This project aimed to characterize the behaviour of wheat elite germplasm at different N regimes. Added to that, our study took place when the BreedWheat project (2011-2019) was testing a similar panel in similar environmental conditions. Thus, results would be easily tested on an independent dataset.

Finally, NUE became an economic, political, and research issue and genetic and genomic resources experienced a burst in bread wheat. This favourable context led to a PhD proposal. Discussions were first engaged between Biogemma and me, after I finished my Master internship in a Limagrain wheat breeding station (Verneuil l'Etang, France). Biogemma was interested in the genetic of NUE in wheat, while I was mainly interested in quantitative genetic methods. The GDEC (Genetic, Diversity, and Ecophysiology of Cereals) was also associated to this project. Indeed, the GDEC is a department of INRA-UBP (French National Institute for Agricultural Research, University Blaise Pascal) which is a major pole of research on cereals. Moreover, its close location facilitates interactions. The Head of the "genetic and genomic of cereals" research group at Biogemma Sébastien Praud, directed this PhD thesis. It was also necessary to have aboard a wheat agronomist specialized in N. Jacques Thus, Le Gouis (GDEC) co-supervised this work. This particular situation of a private/public PhD thesis is governed by a CIFRE contract (industrial contract of formation through research) signed with the ANRT (French National Institut of Research and Technology) which subsidizes private companies hosting PhD students. A collaboration was also set with José Crossa and the CIMMYT (International Maize and Wheat Improvement Center) which provided visiting student with facilities for six months.

RESEARCH STRATEGY

NUE needs to be improved in wheat. This major topic can be addressed from different angles: research on agronomic practices, fertiliser chemistries or genetic improvements. In agreement with Biogemma and GDEC expertises and motivations, we focused on the genetic improvement aspect. The main problematic was: "How can we achieve an efficient breeding for enhanced NUE?"

Breeding is historically achieved through phenotypic selection. Basically, the studied trait is assessed in field trials representing the target environments and on a wide range of progenies from bi-parental crosses

or multi-crosses. Selected lines are self-pollinated and transferred to the next year of trial. After 7-8 cycles of this selection and genetic fixation, a few "fixed" lines are sent to the national registration trials. Nowadays, fixation cycles tend to be shortened by the use of doubled haploids. This selection can also be combined with a selection based on specific genotypic information. This is classically named "marker-assisted selection" (MAS). Genomic selection (GS) is an extension of MAS in which all the genotypic information is used at the same time. Therefore, we will include GS in MAS methods.

Before defining the breeding strategy, for both phenotypic selection and MAS, we need to answer several questions (Table 1) addressing these three inseparable topics: (i) the environments in which lines will be selected and in which varieties will be cultivated, (ii) the germplasm used in breeding and (iii) the targeted traits. Thus, the questions we adressed during this PhD thesis (Table 1) were mostly discussed regarding these three topics. For example, after a review of NUE in wheat, we were interested in analysing the variance of NUE-related traits, keeping in mind that these results depend on the tested environments, germplasm, and traits.

Table 1: Objectives of research and part of the manuscript concerned.

Questions	Part
What is the state of the art of NUE in wheat?	I
What is the variance of NUE-related traits? Is phenotypic selection possible? In which environment? Is it linked to the past breeding efforts?	II
Does it allow us powerful association mapping studies? How can we find genes involved in NUE-related traits genetic determinisms?	II, III
In varieties breeding values (additive or epistatic effects)?	III, IV
In varieties responses to environmental stresses? Which genotypic information should we use in MAS?	IV

We started our work by an analysis of phenotypic variance as the exploitation of this variance is the basis of all our approaches. Indeed, it influences efficiency of phenotypic selection. It also influences efficiency of linkage disequilibrium mapping methods which addresses the specific question of phenotype/genotype associations in MAS. We worked on an historical panel. Thus, we also assessed past breeding progresses. Then, we tried to relate phenotypic and genotypic variance in order to dissect the genetic determinisms of NUE-related traits and to improve MAS methods. Added to that, our work was an opportunity to develop or improve such methods. Thus, across the different questions listed in Table 1, methodological aspects took a significant part of my research.

Two papers were already published and three are ready to be submitted or under reviewing process. Regarding the number of these publications and their complementarity, we found it appropriate to present to the Jury a thesis under the form of articles. Consequently, this manuscript is presented as a compilation of these papers linked by more general discussions and/or further investigations. In each sections, authors' contributions are explicitly described. I hope that you will appreciate what you read. Fabien

REFERENCES

Arvalis institut du végétal : http://agriculture-de-conservation.com/sites/agriculture-de-conservation.com/IMG/pdf/couts_production4869421071029159328.pdf

Biogemma: http://www.biogemma.com/

BreedWheat project: http://www.breedwheat.fr/

Choulet F, Alberti A, Theil S, Glover N, Barbe V, et al. (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science DOI: 10.1126/science.1249721

Eversole K, Feuillet C, Mayer KFX, Rogers J (2014) Slicing the wheat genome. Science 345:285-287

FAO (2011) Current world fertilizer trends and outlook to 2015. ftp://ftp.fao.org/ag/agp/docs/cwfto15.pdf

FAO (2012) World agriculture towards 203/2050, the 2012 revision.. http://www.fao.org/docrep/016/ap106e/ap106e.pdf

Fertilizer Europe: http://www.fertilizerseurope.com/fileadmin/user_upload/publications/communications/2012_Overview-V7.pdf

Internaional Wheat Genome Sequencing Consortium: http://www.wheatgenome.org

Goulding KWT (2004). Minimising losses of nitrogen from UK agriculture. J R Agric Soc Engl 165:1-11

Mayer KF, Martis M, Hedley PE, Simkova H, Liu H et al. (2011) Unlocking the barley genome by chromosomal and comparative genomics. Pant Cell 23:1249-1263

Pathak RR, Lochab S, Raghuram N (2011) Plant Systems | Improving Plant Nitrogen-Use Efficiency. In: Murray Moo-Young (ed.), Comprehensive Biotechnology, 2nd edn. Elsevier, pp 209-218

ProtNblé project: http://www.agence-nationale-recherche.fr/?Projet=ANR-06-GPLA-0016

Wang S, Debbie W, Forrest K, Allen A, et al. (2014) Characterization of polyploidy wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotech J, 12:787-796.

PART I: THE REVIEW

This review is a command of Plant Breeding
Editor: Franck Ordon
Estimated submission: April 2015

BREEDING FOR INCREASED NITROGEN USE EFFICIENCY: A REVIEW FOR WHEAT (*T. AESTIVUM* L.)

Fabien Cormier[1] • John Foulkes[2] • Bertrand Hirel[3] • David Gouache[4] • Yvan Moenne-Loccoz[5] • Jacques Le Gouis[6] ✉

[1]Biogemma, Centre de recherche de Chappes, Route d'Ennezat CS90126, 63720 Chappes, France. [2]Division of Plant and Crop Sciences, The University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, United-Kingdom. [3]Institut Jean-Pierre Bourgin, Département Adaptation des Plantes à l'Environnement, INRA, RD10, 78026 Versailles Cedex, France. [4]Arvalis Institut du Végétal, Station Expérimentale, 91720 Boigneville, France. [5]CNRS, UMR 5557, Ecologie Microbienne, Villeurbanne, France. [6]INRA, UBP UMR 1095, Génétique, Diversité et Ecophysiologie des Céréales, 24 Avenue des Landais, 63177 Aubière Cedex, France.
✉ jacques.legouis@clermont.inra.fr

ABSTRACT: Nitrogen fertiliser is the most use nutrient in modern agriculture and represents significant environmental and production costs. In the meantime, the demand for grain increase and production per area has to increase. In this context, breeding for an efficient use of nitrogen became a major breeding objective. In wheat, nitrogen is required to maintain a photosynthetically active canopy ensuring grain yield and to produce storage protein in the grain hence end-use quality. In different situations of nitrogen management, genetic, metabolic and physiological factor influencing nitrogen uptake and utilization are reviewed. Their implications in breeding are discussed.

DEFINITION OF NUE AND RATIONALE FOR ITS IMPROVEMENT

The concept of nitrogen use efficiency (NUE) has been widely used to characterize plant behaviour regarding different levels of nitrogen (N) availability. It is important to distinguish the concept of NUE and the NUE as a phenotypic trait.

Several definition and evaluation methods have been suggested of which some of them are actually named "nitrogen use efficiency" (reviews in Good et al. 2004, Fageria et al. 2008). Moll et al. (1982) defined the most widespread NUE trait definition, at least among breeders, computed as the grain weight divided by the total N available to plant, and separated it into two components:

$$NUE_{Moll} = NUpE \times NUtE$$

with NUpE the N uptake efficiency calculated as the N in plant at harvest divided by the available N in soil, and NUtE the utilization efficiency calculated as the grain dry mass divided by the total amount of N in plant at harvest. Then, to compute these values when comparing different genotypes, there are two main issues: (i) the complex estimation of N available to crop, and (ii) the estimation of the total amount of N in the plant.

N available to crop results from residual N before sowing, aerial N deposition, mineralization, and the actual availability of applied N. Estimation of these components is complex and an often used proxy has been the amount of applied mineral N fertiliser summed to an estimation of residual N in soil.

Bingham et al. (2012) on 15 barley genotypes compared different methods to estimate available N. The first one was independent to the genotype and used only residual soil N after winter and applied N fertiliser. The two others were dependent to the genotype and required a control without N fertilisation (N^0). Available N for the fertilized treatment (N^T) is then estimated either (i) by adding the total plant N at harvest for N^0 to the applied N fertiliser or (ii) by adding soil N at harvest to (i). Bingham et al. (2012) showed that genotype rankings are very similar between the three methods and that the simplest method can be used to start with.

Although, as discussed in Cormier et al. (2013), these can lead to overestimation of NUE in low N situations and underestimation in high N situations making comparison and/or joint analyses of different studies difficult. Within a large collection of genotypes, Cormier et al. (2013) suggested estimating available N from the distribution of the total plant N at harvest. They proposed to use N absorbed by the top 5% genotypes as an estimation of N that was available to the whole series.

To estimate the total amount of N in the plant, usually only the aerial parts are sampled. Not taking into account N in the roots would increase NutE and decrease NupE. However, measuring the quantity of roots N (in the first 30 cm of soil layer) of a set of cultivars grown at two N levels, Allard et al. (2013) showed that only a small fraction of total N is partitioned to the roots (about 4 % or 10 kg ha^{-1} at harvest). Here again the genotype rankings were very similar with or without root N.

Looking at the successes and debates that agitated other scientific community may help to improve the approaches on NUE. Ecologists developed another decomposition of NUE. Originally called "nitrogen utility", Hirose (1971) defined it as the flux ratio of dry mass productivity for a unit of N taken up from

the soil. Berendse and Aerts (1987) suggested dividing it into two components to make it biologically meaningful in a context of perennial species in a steady-state system (*i.e.* annual biomass production = annual biomass loss; annual N uptake = annual N loss). Thus, NUE was defined as the product of the nitrogen productivity rate (NP; dry mass growth per unit plant N) and the mean time residence of N (MRT). Later, Hirose (2011) revisited this definition and specified how it should be calculated to make it also suitable for non-steady state system such as annual crops.

Compare to Moll et al. (1982), this definition has the interest to deliver a dynamic vision of NUE directly related to photosynthetic activity along the plant cycle. Nevertheless, it only focuses on utilization and plant efficiency to extract N from the soil is not taken into account. However, in annual crops, this is an important parameter to consider as substantial amounts of N fertiliser are applied, implying environment and economic issues.

In a similar way, in the water use efficiency (WUE) community, it has been explicitly decided not to account for plant available water, and the focus has been on viewing yield as the final objective through Passioura's (1977) seminal equation:

$$GY = WU \times WUE \times HI$$

with WU the water use (mm transpired), WUE the water use efficiency (kg aboveground DM / mm transpired) and HI the harvest index (kg grain / kg above-ground dry matter).

Paralleling to NUE_{Moll} formalization, NUtE would then be equivalent to WUE × HI. NUpE would be an equivalent to WU divided by plant available water. The approach could be taken further by simply targeting nitrogen use (NU) as kg N absorbed by the plant instead of NUpE; in much the same way that WU is seen as (arguably) the most important target in improving water response (Blum, 2009). This would also avoid dividing an already rather imprecise variable (NU) by an even more imprecise one (available N).

Yet, environmental and economic issues are different in NUE where minimizing the loss of fertiliser applied (*i.e.* by leaching) and maximizing N uptake for increasing grain protein concentration lead to focus also on NUpE. Moreover, not to account for N available to crop imply to use genotypes dependent methods (*i.e.* repeated controls) to compare varieties behaviour between different stress intensities or to characterize genotypes × stress interaction, leading to confounding effects.

Critiques of the initial WUE equation have heavily contributed to identify and prioritize approaches and traits. The first has been to recognize that the three terms of the equation are clearly not independent (Blum, 2009; Tardieu, 2013). Typically, as WU increases, WUE decreases because WU scales to biomass (Blum, 2009), as does N absorption (Sadras and Lemaire, 2014; Lemaire et al. 2007). Consequently, an excessively narrow focus on WUE may prove counterproductive (Blum, 2009). Although, the underlying physiological reasons for this are very different between nitrogen and water, framing the nitrogen community in much the same way as the water community would help in placing the focus NU and on systematically accounting for total biomass when evaluating NU, as advocated for by Sadras and Lemaire (2014). As in water and ecologist communities, research on NUE can also

be disconnected of the Moll et al. (1982) NUE definition and focus on a dynamic approach. Indeed, NUpE and NUtE are calculated at the end of the crop cycle. Although total N in plant varies during the cropping season and have a critical interaction with HI: once grains are growing, they become a N sink, and growers, breeders, and the wheat industry as a whole must manage the contradictory objective of high yields and high protein contents (Feil, 1997; Jeuffroy et al. 2002; Oury & Godin, 2007).

First of all, pre-anthesis and post-anthesis phases should be clearly separated. Regarding the post-anthesis phase, the grain protein deviation (GPD; deviation from the yield-protein regression) criterion suggested by Monaghan et al. (2001) and Oury & Godin (2007) allows to specifically breed for high protein without the associated yield penalty. Bogard et al.'s (2010) analysis of GPD showed that this metric was tightly related to another deviation: that between pre-anthesis N uptake and post-anthesis N uptake meaning the obvious: crops that are both high yielding and high in protein content absorb large quantities of nitrogen. In other words, Bogard et al.'s (2010) analysis places NU as a key factor without focusing on NUpE. Looking now to the pre-anthesis phase has the advantage of not having to deal with the yield-protein trade-off. Studying N impacts on yield, grain number per area can become the criterion to target instead of yield. Indeed, it allows to get rid of kernel weight elaboration, which occurs post-anthesis and as suggested by Meynard (1987), at least in western European situations, N will essentially have an impact on grain number per area, and kernel weight will often add noise due to other stresses. This would mean that HI would essentially be replaced by an FI (fertility index). This implies complex phenotyping although it may allow a better characterization of N response regarding the phenologic stage.

NUE has been the subject of a wealth of literature and underpinning projects for its improvement. However, there seems to be consensus on the need to increase progresses on NUE in breeding. To the best of our knowledge, NUE has not been the target of dedicated breeding improvement. Rather, it has been improved through indirect selection for yield, in those environments targeted by breeding programs. Sadras and Richards (2014) have suggested that indirect selection for yield serve as a benchmark for any alternative approach. Several studies have evaluated a posteriori breeding improvement in NUE (Ortiz-Monasterio et al. 1997a; Guarda et al. 2004; Muurinen et al. 2006; Cormier et al. 2013). Taking the case of France as an example, Cormier et al. (2013) quantified NUE_{Moll} improvement at 0.13 kg DM kg^{-1} N $year^{-1}$. Supposing an average French yield of 7 t / ha, and assuming a reference NUE value between 37.8 kg DM kg^{-1} N (Cormier et al. 2013) and 33.3 kg DM kg^{-1} N (average value for wheat used in French balance sheet N recommendation methods; Meynard, 1987), this equates to a saving of around 6-8 kg N ha^{-1} after 10 years of genetic improvement. From this economic standpoint, the variations in (fertiliser N / grain price) ratios essentially determine the quantity of N applied. The impacts of this volatility on on-farm NUE and required N savings can be translated into two examples. Sylvester-Bradley and Kindred (2009) showed that this price ratio in the past 10 years has varied from 3 to 9 (Sylvester-Bradley and Kindred, 2009) leading to a necessity to increase NUE from

23.8 to 28.6 kg DM kg^{-1} N. Thus, it would require almost 40 years of breeding progress to compensate the variations generated by volatile N : grain price ratios. Over the same period, 16% of the total observed volatility was a variation of N : grain price ratio from 5 to 6 (Cohan, 2009) leading to a necessity to economize 6-7 kg N ha^{-1} corresponding as previously mentioned to 10 years of improvement.

Overall, this leads us to conclude that breeding needs to tackle NUE more efficiently than it has been doing at the current rate.

TRAITS INFLUENCING N-UPTAKE EFFICIENCY

Root size and morphology

Nitrate is readily leached down the soil profile and consequently the primary root traits to improve for enhanced N capture include rooting depth and rooting density, especially for post-anthesis N uptake (Foulkes et al. 2009). A deeper relative distribution of roots could comprise part of an ideotype to maximize N capture and further improvements in root architecture could focus on root proliferation at depth in wheat (Carvalho and Foulkes, 2011). Indeed, root length density (root length per unit volume of soil) is often below a critical threshold for potential nitrate capture of around 1 cm cm^{-3} (Barraclough et al. 1989; Gregory and Brown, 1989) at lower depths in the rooting profile (Ford et al. 2006; Reynolds et al. 2007).

Genetic variation in root system size has been widely reported in wheat (e.g. O'Toole and Bland, 1987; Hoad et al. 2001; Ehdaie and Waines, 2003; Ford et al. 2006), but root distribution varies strongly with soil characteristics, nutrient availability and mechanical impedance. In wheat, the use of synthetic wheat derivatives, incorporating genes from the diploid wild species *Triticum tauschii* (D genome) with roots distributed relatively deeper (Reynolds et al. 2007) may help in the development of cultivars with relatively deeper rooting systems. In addition, the wheat-rye translocation in 'Kavkaz' for the short arm of chromosome 1 (1RS) has been observed to have increased root biomass at depth (Ehdaie et al. 2003) and tall landraces from China and Iran had larger root biomass than semi-dwarf cultivars descended from CIMMYT breeding material (Ehdaie et al. 1991; Ehdaie and Waines, 1993, 1997; Ehdaie, 1995). It may also be possible to increase root length density at depth without extra carbon input by modifying specific root length (root length per root biomass; Carvalho et al. 2014). Although it is well established that plants respond to N deficiency by increasing the ratio of root biomass on total plant biomass (root dry weight ratio; RDWR) due to the functional equilibrium between the growth of the root and shoot (Barraclough et al. 1989; Dreccer et al. 2000; Robinson et al. 2001), there are to date no reports of genetic variation in the dynamic responses of RDWR to N supply.

Direct selection for root system architecture traits (length, biomass, density, lateral root dispersion) has been associated with improved water and/or nutrient uptake in wheat (Hurd, 1964), upland rice (Price et al. 2002) and maize (Lynch, 2007). Indirect selection for lower canopy temperatures might also be taken as an indication of a greater root uptake capacity, but higher stomatal conductance would produce a similar signal

(Reynolds et al. 2009). Root hairs provide another potential mechanism to maximize N capture and two genes for root hair elongation, RTH1 and RTH3, have been identified in maize (Hochholdinger and Tuberosa, 2009). Root architecture and root function are likely to be multigenic and hence much more difficult to select for (Hall and Richards, 2013). Therefore, breeding for root characteristics has seldom been implemented to date, principally because of the difficulties of scoring root phenotypes directly and the absence of suitable proxy measurements. Nevertheless, marker-assisted selection may be especially useful to pyramid multiple traits, such as root angle, root length, root weight and root to shoot ratio, which are associated with main effect quantitative trait locus (QTL) in wheat (Hamada et al. 2012; Sharma et al. 2011; Bai et al. 2013), even if a better understanding of the biology of these traits and the potential synergies and trade-offs between traits is required (Lynch et al. 2007). For example, the expression of length and density of root hairs may be synergistic (Ma et al. 2001) and there may be antagonistic interactions between biomass allocation to different root classes due to assimilate competition (Walk et al. 2006).

Root N transporter systems

In most countries, the commercial mineral forms of N commonly applied to crops growing on cultivated soils, are anhydrous ammonia, urea, ammonium sulphate and ammonium nitrate (Robertson and Vitousek, 2009; Andrews et al. 2013). In addition, farmyard manure is also able to supply a considerable amount of N fertilisation (Hooda et al. 2000; Körschens et al. 2013). Mineral N fertilisers are particularly soluble for easy assimilation by crops. Both urea and ammonia are converted to nitrate (NO_3^-) at different rates depending on the nature of the soil and of the climatic conditions (Jarvis et al. 2011). Thus, NO_3^- is the main source of N for most crop species, whether inorganic or organic N is provided to the plant (Nasholm et al. 2009; Gioseffi et al. 2012).

Ammonium (NH_4^+) is the ultimate form of inorganic N available to the plant. Most of the NH_4^+ incorporated by the plant into organic molecules originates from NO_3^- reduction, although metabolic pathways such as photorespiration, phenylpropanoid metabolism, utilization of N transport compounds and amino acids catabolism can generated NH_4^+ (Lea and Miflin, 2011). Nevertheless, despite active nitrification mechanisms by soil microorganisms, substantial amounts of ammonium (NH_4^+) can remain, but the NH_4^+ concentration is generally ten times lower compared to that of NO_3^- in cultivated soil (Nieder et al. 2011). Both NO_3^- and NH_4^+ enter the root apoplast by diffusion or mass flow (Crawford and Glass, 1998). Then, there are taken up via an active transport system by means of proteins termed high and low affinity transporters located in the root cell plasma membrane (Loqué and von Wirén, 2004; Glass 2009; Dechorgnat et al. 2011).

In higher plants, there are basically three different NO_3^- transport systems that operate depending on the NO_3^- concentration in the surrounding root environment. The first is an inducible high affinity transport system (iHATS) that is induced in the presence of low concentration of NO_3^- in the range of 1 to 200 µM depending on the plant species examined (Pace and McClure, 1986; Sidiqui et al. 1990). In wheat it was reported that the iHATS has

a K_m value of approximately 27 µM and requires 10 hours for full induction by NO_3^- (Goyal and Huffaker, 1986). The second is a constitutively expressed high affinity transport system (cHATS) that is present even in the absence of NO_3^-. Both systems exhibit a typical Michaelis-Menten saturation profile when the external NO_3^- concentration reaches a certain threshold. The third is represented by a non-saturable low affinity transport system (LATS) that dominates when NO_3^- in the external medium exceeds 250 µM operating in the concentration range of 0.5-1 mM (Sidiqui et al. 1990; von Wirén et al. 1997). Recent studies of NO_3^- channels of transporters showed that they can also play versatile roles in sensing NO_3^-, in plant development, in pathogen defence and in stress response (Wang et al. 2012). Although NH_4^+ ions can be passively taken up by plant roots, different transport root NH_4^+ transporters system (Ludewig et al. 2007) allow the direct uptake of NH_4^+ ions and operate in a wide range of NH_4^+ concentrations (Loqué and von Wirén N. 2004). However, it is likely that in agricultural soils NH_4^+ uptake operates mainly through the low affinity transport system (LATS), which is part of the NH_4^+ permeases in the Ammonium Transporter / Methylammonium Permeases / Rhesus (AMT / MEP / Rh) family (von Wirén and Merrick, 2004). The K_m values for NH_4^+ influx in different species ranges between 1 to 200 µM (Bradley and Morris, 1991; Wang et al. 1993), fitting with the average NH_4^+ soil concentration which rarely rises beyond 50 µM (Marshner, 1995). In wheat, it was reported that the iHATS has a K_m value of approximately 50 µM and requires six hours for full induction by NH_4^+ (Goyal and Huffaker, 1986).

Nitrate (NO_3^-) transporters in higher plants are represented by two main families of genes namely the NRT1 PTR (Nitrate Transporter, Peptide Transporter) Family (NPF), which now regroups the previous NRT1 / PTR genes, and NRT2 family also called the Major Facilitator Superfamily (MFS; Léran et al. 2014). An excellent review describing the different members of the NO_3^- and NH_4^+ transporters and the regulatory mechanisms affecting root N uptake systems, especially on the model species Arabidopsis, has recently been published by Nacry et al. (2013). This review emphasizes that expression and activity of most N uptake systems are regulated both by the concentration of their substrate and by a systemic feedback control of metabolites representative of the whole plant N status. In cereals in general and wheat in particular, there is far less information on the root NO_3^- and NH_4^+ transport systems and their regulation. This is mainly because most of the pioneer work was conducted using the model plant Arabidopsis, due to the ease of obtaining mutants and transgenic plants altered in the expression of the different NO_3^- and NH_4^+ transporters (Miller and Smith, 1996; von Wirén and Merrick, 2004; Miller et al. 2007; Garnett et al. 2009; Xu et al. 2012). Gene structure and phylogeny of high or low affinity transport systems have been studied in a number of grasses including rice, maize, sorghum, *Brachypodium* and wheat (Plett et al. 2010; Yin et al. 2007; Girin et al. 2014).

Very recently, a comprehensive overview of the complex phylogeny and gene expression patterns of 16 members of the NPF family in wheat has been published (Buchner and Hawkesford, 2014). This study highlighted the complex pattern of expression of the nitrate transporters, mainly due to the

presence of multiple co-orthologous genes that are differentially expressed according to the plant tissue, NO_3^- availability and to leaf senescence during the N assimilation and N remobilisation processes. Earlier studies have also demonstrated that in the wheat NO_3^- HATS system, five genes are induced by abscisic acid when NO_3^- is not present. In contrast to the inhibitory effect of glutamine generally observed in other species, glutamine was able to induce the expression of NRT2 genes in the absence of NO_3^- (Cai et al. 2006). In addition, it also has to be considered that under agronomic conditions, both efficiency and the regulation of NO_3^- uptake systems may be enhanced by the presence of mycorrhizal associations (Hawkins et al. 2001), humic substances (Cacco et al. 2000), allelopathic compounds such as coumarin (Abenavoli et al. 2001) and root growth promoting bacteria (Mantelin and Touraine, 2004) or inhibited when the CO_2 concentration is rising in the atmosphere (Bloom et al. 2014). Therefore such environmental interactions, together with the capacity of the plant to capture and transport NO_3^- or NH_4^+ must be taken into account, particularly when studying the genetic basic of inorganic N uptake during the pre- and post-anthesis period.

This implies that, in combination with modelling approaches (Bertheloot et al. 2011), further research is required to obtain an understanding of the regulation of the NO_3^- and NH_4^+, HATS and LATS throughout the entire plant developmental process (Kong et al. 2013). It will also be necessary to evaluate the contribution of direct NH_4^+ uptake to the wheat N economy, as in wheat (Causin and Barneix, 1993; SØgaard et al. 2009) and other cereals such as maize (Gu et al. 2013) and rice (Gaur et al. 2012), the available information on the NH_4^+ transport systems both at the molecular and physiological levels, remains fragmentary. However, for wheat that preferentially uses NO_3^- instead of NH_4^+ as the main N source, an increase in NH_4^+ uptake may not be beneficial to the plant when the ion is applied to the soil (Angus et al. 2014).

Another field of investigation is the use of urea as a synthetic fertiliser in conventional agriculture (Andrews et al. 2013; Karamos et al. 2014). Indeed, to date, urea is mainly used as a source of N fertiliser (through soil mineralization after application) and the contribution of plant urea uptake and metabolism in a physiological and agricultural context has not been thoroughly investigated. Nevertheless, it is well known that plants possess leaf and root transporters to absorb urea as an intact molecule, and can hydrolyse and use it very efficiently (Witte 2011). Two distinct transport processes for urea have been identified in rice exhibiting a linear and a Michaelis-Menten kinetics with an affinity for NH_4^+ ranging from 40 to 1000 µM (Wang et al. 2012). Moreover, encouragingly, when a rice urea transporter was overexpressed in Arabidopsis a positive effect was observed both on urea uptake at low concentration and on plant growth (Wang et al. 2012). In wheat, compared to other inorganic N sources, urea uptake was very low. Moreover, its kinetics of uptake was difficult to measure (Criddle et al. 1988). However, in some cases when applied at an optimum timing after anthesis, an increase in grain protein content or yield has been observed (Gooding and Davies, 1992; Rawluk et al. 2000). More recently, it has been shown that in spring wheat seed yield and N uptake were generally greater with polymer coated urea than urea alone (Malhi and Lemke, 2013).

Even if the efficiency of foliar application of urea in wheat and other cereals remains questionable, it is attractive in terms of environmental benefit. More research is thus required both at physiological and molecular levels.

Interaction with micro-organisms

Plant roots, including those of wheat, release a variety of organic substrates (*e.g.* organic acids, and sugars), exudates and other rhizodeposits (Nguyen 2003). This creates a particular fraction of soil in contact with roots named rhizosphere and favourable to microorganisms development. Plant rhizosphere is largely colonized by soil microorganisms, at levels of typically 10^8 to 10^9 bacteria per gram of rhizosphere soil and 1 to 1.5 m of fungal filaments per cm^2 of root surface (Moënne-Loccoz et al. 2014). This microbial community contains a broad range of taxa differing from bulk soil community due to the selective effects of roots (Buée et al. 2009). Some of them, including pathogens as well as non-pathogenic microorganisms, may enter roots and reside within intercellular space or even within plant cells (Behl et al. 2012, Moënne-Loccoz et al. 2014). This also occurs in wheat (Germida and Siciliano 2001).

The composition and physiological activities of root-associated microbial communities is influenced by many factors, such as soil characteristics, farming practices, climatic conditions, and wheat genotypes (Mazzola et al. 2004). Indeed, rhizodeposition can differ between wheat cultivars (Wu et al. 2001), which can lead to differences in various aspects of the rhizosphere microbial ecology (Germida and Siciliano 2001). Therefore, it would be of prime interest to develop breeding strategies tailored both to suppress root pathogens and promote root colonization by plant-beneficial microbial partners (Lammerts van Bueren et al. 2011), especially those with the potential to enhance (i) N availability in the rhizosphere, (ii) root system and architecture, (iii) systemic plant metabolism and (iv) microbial phytoprotection (Fig. 1). This is all the more relevant since breeding is typically carried out under optimal conditions, thus plant traits involved in plant growth-promoting rhizobacteria interaction may have been neglected (den Herder et al. 2010).

Soil microorganisms in the rhizosphere are major players in the availability of N for plant roots (Richardson et al. 2009). On one hand, N availability for roots may be reduced by microbial competition. Indeed, various soil bacteria and fungi use ammonium and nitrate as N sources (Nelson and Mele 2006) and/or transform nitrate to gaseous N by denitrification (Herold et al. 2012). Nevertheless, plants can limit denitrification by releasing inhibitory secondary metabolites (Bardon et al. 2014), but so far this property is not documented in cultivated cereals. However, attempts are currently made to introduce into wheat a chromosome of *Leymus racemosus*, a wild relative of wheat, containing the ability for biological nitrification inhibition (Subbarao et al. 2007; Ortiz et al. 2008).

On the other hand, N availability for roots is enhanced by microbial mineralisation of organic N yielding ammonium in the rhizosphere. This entails proliferation of bacterial and fungal decomposers, as well as protozoan predators (Bonkowski 2004) and mycorrhizal fungi (Atul-Nayyar et al. 2009). In wheat, this priming effect reaches higher levels at the flowering stage (Cheng et al. 2003) and root

colonization by mycorrhizal fungi as well as positive mycorrhizal effects on plant nutrition and yield is genotype-dependent (reviewed in Behl et al. 2012).

N availability for roots is also improved by N fixation. Thus, the community of N fixers (functional group) plays a key role for plant N nutrition (Hsu and Buckley 2009). Unlike in legumes, conversion of N_2 into NH_3 in wheat and other cereals does not entail root-nodulating rhizobia but it can be perfomed by other non-nodulating N-fixing bacteria, and part of the N fixed may be acquired by the plant (Behl et al. 2012). N-fixing bacteria occur naturally in soils including in the wheat rhizosphere (Nelson and Mele 2006; Venieraki et al. 2011), and inoculation with N fixers may enhance wheat yield (Kapulnik et al. 1987, Hungria et al. 2010, Behl et al. 2012, Neiverth et al. 2014). Their diversity and activity fluctuate with both plant species (Perin et al. 2006, Reardon et al. 2014) and cultivar (Coelho et al. 2009) including in wheat (Christiansen-Weniger et al. 1992, Manske et al. 2000; Venieraki et al. 2011). For example, N-fixing bacteria *e.g. Azospirillum brasilense* Sp245 have limited potential to improve wheat nutrition (Baldani and Baldani 2005), whereas others *e.g. Klebsiella pneumoniae* 342 can relieve N deficiency and enhance N levels (Iniguez et al. 2004) depending on cultivar (Manske et al. 2000).

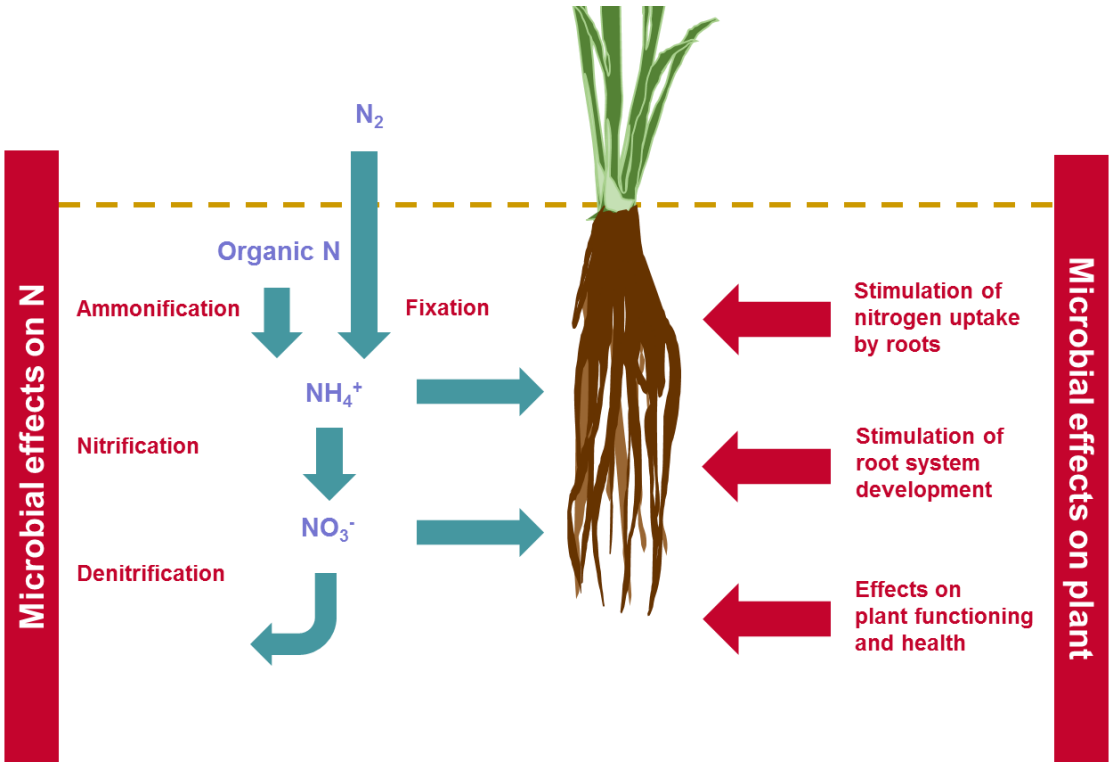


Figure 1: Summary of microbial effects.

Enhanced acquisition of water and mineral nutrients can be expected if the root system colonizes soil more extensively. Under in vitro conditions, wheat inoculation with rhizosphere bacteria may enhance root number and/or length, as well as root hair elongation (Dobbelaere et al. 1999, Combes-Meynet et al. 2011). These inoculation effects on root system architecture and biomass have been

also evidenced in soil-grown wheat (Baldani and Baldani 2005, Veresoglou and Menexes 2010). Indeed, many bacteria and fungi modify root system architecture by manipulating plant hormonal balance, in particular by producing phytohormones such as auxins (Ortíz-Castro et al. 2009), cytokinins (Cassán et al. 2009, Moubayidin et al. 2009), or gibberellins which are produced by several rhizosphere bacteria and fungi (Bottini et al. 2004) including wheat strains (Upadhyay et al. 2009) promoting primary root elongation and lateral root extension. The wheat bacterium *Azospirillum brasilense* Sp245 synthesizes abscisic acid, which modifies lateral root development, and inoculation resulted in higher abscissic acid concentration in *Arabidopsis* (Cohen et al. 2008). Other root-branching signals especially 2,4-diacetylphloroglucinol (Brazelton et al. 2008) and nitric oxide (Creus et al. 2005) may also be implicated, including in wheat (Pothier et al. 2008, Couillerot et al. 2011). Their effects appear to take place via an auxin signal transduction pathway (Brazelton et al. 2008, Molina-Favero et al. 2008). Microbial interference with ethylene metabolism in roots may also be responsible for modifying wheat root system architecture (Upadhyay et al. 2009) by a direct microbial production of ethylene (Graham and Linderman 1980), or a reduction of ethylene concentration in plant roots by the deamination of ethylene precursor 1-aminocyclopropane carboxylic acid (Prigent-Combaret et al. 2008), thereby diminishing ethylene-mediated root growth repression (Glick 2005).

Microorganisms can induce systemic changes in plant physiology. For instance, a wide range of Arabidopsis genes displayed different expression levels upon inoculation with a plant-beneficial *Pseudomonas putida* bacterium (Srivastava et al. 2012). Microbial inoculation may also modify plant proteomic profiles (Mathesius 2009) and metabolomics profiles, both for primary metabolites (including rice shoot contents in amino acids; Curzi et al. 2008) and secondary metabolites in maize (Walker et al. 2012) and wheat (Fester et al. 1999). There are also indications that certain rhizosphere bacteria may directly affect N metabolism in plants. Oil seed rape (*Brassica napus* L.) roots inoculated with *Achromobacter* strain U80417 displayed enhanced net influx rates of NO_3^- (Bertrand et al. 2000), and genes coding for two nitrate transporters (NRT2.5 and NRT2.6) were expressed at higher levels in Arabidopsis upon inoculation with *Phyllobacterium brassicacearum* STM196 (Mantelin et al. 2006). Exposure of tomato to the bacterial metabolite 2,4-diacetylphloroglucinol increased the net root efflux of amino acids (Phillips et al. 2004). In addition, nitrate reductase activity of *Azospirillum brasilense* Sp245 inside roots is thought to contribute to N assimilation of wheat (Baldani and Baldani 2005). However, information is scarce and relevance for wheat remains to be investigated.

A range of root-associated microorganisms promote plant health, by inhibiting root pathogens and/or systemic induction of plant defence mechanisms (Couillerot et al. 2011, Almario et al. 2013). For instance, wheat inoculation with the bacterium *Pseudomonas fluorescens* Q8r1-96 resulted in cultivar-dependent, defence-related transcript accumulation in roots (Maketon et al. 2012). Thus, microbial phytoprotection effects are also important to consider and investigate.

TRAITS INFLUENCING N-UTILIZATION EFFICIENCY

Nitrate assimilation

After being taken up by the roots, nitrate (NO_3^-) is then reduced to nitrite (NO_2^-) in the cytosol through the reaction catalysed by the enzyme nitrate reductase (NR; EC 1.7.1.1) using NADH / NAD(P)H / NADPH as electron donors. The NR enzyme represents the first step in the pathway of NO_3^- assimilation. They are positively regulated by NO_3^- and light at the transcriptional level; and is down regulated at the post-transcriptional level by reversible phosphorylation during the dark period (Kaiser et al. 2011). In hexaploid wheat, two genes encoding NADH-NR have been identified (Boisson et al. 2005). NO_3^- reduction is followed by the reduction of NO_2^- to NH_4^+ catalysed by the enzyme nitrite reductase located in the plastids (NiR; EC 1.7.7.1; Sétif et al. 2009). NiR forms a complex with Ferredoxin that provides electrons for the reduction of NO_3^- to NH_4^+ (Sakakibara et al. 2012). Ammonia (NH_4^+) is then incorporated into the amino acid glutamate through the action of two enzymes. The first reaction catalyzed by enzyme glutamine synthetase (GS; EC 6.3.1.2; Lea and Miflin, 2011) is considered to be the major route facilitating the incorporation of inorganic N into organic molecules in conjunction with the second enzyme glutamate synthase (GOGAT; EC 1.4.7.1; Suzuki and Knaff, 2005), which recycles glutamate and incorporates C skeletons in the form of 2-oxoglutarate into the cycle. The amino acids glutamine and glutamate are then further used as amino group donors to all the other N-containing molecules, notably other amino acids used for storage, transport and protein synthesis and to nucleotides used as basic molecules for RNA and DNA synthesis (Lea and Miflin, 2011; Fig. 2).

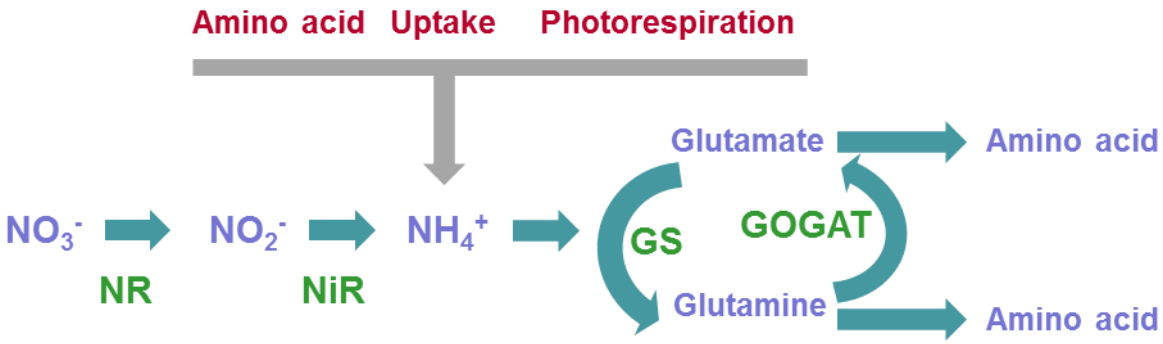


Figure 2: Main N assimilation pathways in wheat.

In higher plants, including wheat, the two enzymes GS and GOGAT are present in the plant in several isoenzymic forms located in different cellular compartments and differentially expressed in a particular organ or cell type according to the developmental stage. The GS enzyme exists as a cytosolic form (GS1) present in a variety of organ and tissues such as roots, leaves, phloem cells and a plastidic form (GS2) localized in the chloroplasts of photosynthetic tissues and the plastids of roots and etiolated tissues. The relative proportions of GS1 and GS2 vary within the organs of the same plant and between plant species, each GS isoform playing a specific role in a given metabolic process, such as

photorespiratory ammonia assimilation, nitrate reduction, N translocation and recycling (Lea and Milfin, 2010). In wheat and other C3 cereals, both at the transcriptional and at enzyme activity levels GS2 predominates throughout the entire plant developmental cycle, although its activity can decrease by half after the flowering period. One GS1 isoenzyme is constitutively expressed in the phloem while the other is generally induced in the cytosol of senescing leaves (Kichey et al. 2005; Christiansen and Gregersen, 2014; Yamaya and Kusano, 2014.). Detailed analyses of gene expression and cellular localization of the different wheat GS isoenzymes were performed in developing and senescing leaves as well as in a number of reproductive tissues (Kichey et al. 2005; Bernard et al. 2008). These studies have highlighted the complex GS isoenzyme pattern of expression not only due to the hexaploid nature of the wheat genome, but also due to the morphological complexity of the leaves. In order to clarify the function of the different GS isoenzymes, a phylogenetic approach was taken, due to the lack of mutants or transgenic plants. This allowed the division of the different genes encoding GS into different classes of biological functions, which were not necessarily conserved between C3 and C4 cereals (Thomsen et al. 2014).

The enzyme GOGAT also exists in two forms that have specific roles during primary N assimilation or N recycling. A ferredoxin-dependent isoenzyme (Fd-GOGAT) is mainly involved, in conjunction with GS2, in the reassimilation of photorespiratory ammonia and a pyridine nucleotide-dependent isoenzyme (NADH-GOGAT; EC 1.4.1.14) is involved in the synthesis of glutamate both in photosynthetic and non-photosynthetic organs or tissues, to sustain plant growth and development (Lea and Miflin, 2011).

Glutamate can also be generated by the incorporate of ammonia into 2-oxoglutarate by the enzyme glutamate dehydrogenase (GDH; EC 1.4.1.2; Lea and Miflin, 2011). However, a number of experiments using ^{15}N-labelling techniques and mutants deficient in GS and GOGAT have demonstrated that over 95 % of the ammonia made available to the plant is assimilated via the GS / GOGAT pathway (Lea and Miflin, 2011). Later on, it was clearly demonstrated that GDH operates in the direction of glutamate deamination to provide organic acids, notably when the root and leaf cells are carbon-limited (Labboun et al. 2009; Fontaine et al. 2012). Recently, the hypothesis that GDH plays an important role in controlling not only glutamate homeostasis (Forde and Lea, 2007; Labboun et al. 2009), but also the level of downstream and upstream carbon and N metabolites through the changes in the hetero-hexameric structure of the enzyme, has been put forward (Tercé-Laforgue et al. 2013). This function, which may also have a signalling role at the interface of C and N metabolism, may be of importance when there is a shortage of C under stress conditions or during certain phases of plant growth and development. Moreover, both transgenic studies performed on a number of model and crop species (Tercé-Laforgue et al. 2013) and quantitative genetic approaches performed on maize (Dubois et al. 2003) and wheat (Fontaine et al. 2009), strongly suggest that the reaction catalysed by NAD(H)-GDH is of major importance in the control of plant growth and productivity. Further research is thus required to validate the function of GDH in crops such as wheat.

Over the last two decades, our knowledge of the various pathways involved in the synthesis of the amino acids, particularly those derived from glutamate and glutamine, has been increased through the use of mutant and transgenic plants in which amino acid biosynthesis has been altered. Amino acid biosynthesis is of major importance for cereal growth and productivity (Howarth et al. 2008). There are excellent reviews describing extensively the current knowledge on this complex pathways and its regulation (Lea and Azevedo, 2007), therefore we will not cover it in this review.

Leaf and canopy photosynthesis per unit N

Up to 75% of N in wheat leaves is located in mesophyll cells, mainly as the chloroplastidic enzyme Rubisco, and is involved in photosynthetic processes (Evans, 1983). Thus, responses in N-limited crops often include reductions in total leaf area, leaf expansion and duration, leaf N and chlorophyll content, leaf stomatal conductance, and photosynthesis per unit leaf area (Sylvester-Bradley et al. 1990; Monneveux et al. 2005). These responses reduce radiation interception and radiation-use efficiency (above-ground biomass per unit radiation interception; RUE) and hence biomass (Foulkes et al. 2009b) and yield. Canopy and leaf processes affecting photosynthesis per unit N uptake include: (i) radiation interception per unit N uptake, (ii) optimizing vertical N distribution in relation to light in the canopy and (iii) leaf photosynthesis per unit leaf N.

For 95 % radiation interception assuming a light extinction coefficient (K) value of 0.5, a green area index (green canopy area per unit ground area; GAI) of 6 is required. Indeed,

$$K = - \ln (I / Io) / L$$

where Io is the incident radiation and I is the amount of radiation not intercepted by a canopy having a GAI = L.

At anthesis, modern wheat cultivars produce canopies with GAI values in the region of 6 hence achieve full interception at this stage (*e.g.* Moreau et al. 2012; Gaju et al. 2014). The only realistic way to increase fractional interception in the pre-anthesis phase is by increasing fractional interception at the start of the stem-elongation phase. However, it is in the region of 60-70 % in wheat (Shearman et al. 2005; Moreau et al. 2012). Thus, only marginal improvement seems possible. Physiological avenues for increasing fractional interception specifically under low N supply may include: increased specific leaf N area (leaf area per unit leaf N; SLN) or/and higher light extinction coefficient. Genetic variation in SLN has been associated with embryo size (Lopez-Castaneda et al. 1996) and earlier canopy closure (Rebetzke & Richards, 1999). The light extinction coefficient is mainly influenced by leaf angle. For modern wheat cultivars is approximately 0.55 for photosynthetically active radiation (Thorne et al. 1988; Abbate et al. 1998; Moreau et al. 2012). These values are associated with semi-erect to erect leaf angles which help to reduce light saturation in the upper canopy leaves boosting RUE. A higher value of K seems unlikely to be desirable due to the trade-off with RUE. Although desirable, more prostrate leaves during early vegetative growth and more upright leaves during later vegetative growth may be difficult to achieve in practice. In summary, although genetic gains in radiation interception per

unit N uptake may be possible during stem elongation they seem likely to be small.

N distribution in canopies in relation to light attenuation also affects photosynthesis per unit N uptake. Considering that the leaf N gradient is "optimal" in accordance with the "optimization theory" (Field, 1983; Hirose and Werger, 1987; Anten et al. 1995; Moreau et al. 2012), theoretical studies indicated that leaf N maximizes canopy photosynthesis when it parallels the light gradient, *i.e.* when the light (K_L) and N (K_N) extinction coefficients are equal. In wheat, observed N gradients are generally less steep than predicted with the optimization theory, however do demonstrate that SLN follows an exponential gradient with vertical depth in the canopy (Critchley, 2001; Pask, 2009; Moreau et al. 2012). Possible reasons for this discrepancy have been discussed in detail by Kull (2002). There is relatively little information on genetic diversity in the vertical distribution of N in relation to light in the canopy. Nevertheless, Berteloot et al. (2008) demonstrated with two French winter wheat cultivars (Apache and Isengrain) that the vertical distribution of N at anthesis was close to the optimum, as defined in the optimization theory, and only differed significantly at the end of grain filling. Similarly, genetic differences were not found for five spring wheat genotypes grown in the Netherlands (Bindraban, 1999). Moreau et al. (2012) analysed the vertical distribution of leaf N and light at anthesis for 16 wheat cultivars experimented in field trials in France and the UK in two seasons under two N levels. The N extinction coefficient with respect to light (K_N/K_L) varied with N supply and cultivar. A scaling relationship was observed between K_N:K_L and the size of the canopy for all the cultivars in the different environmental conditions. Interestingly, the scaling coefficient of the K_N:K_L - green area index relationship differed among cultivars, suggesting that cultivars could be more or less adapted to low N environments.

Photosynthesis rate per unit N affects NUtE. In C3 cereals such as wheat, the net light-saturated rate of leaf photosynthesis (Amax) typically increases to 20-30 μmol CO_2 m^{-2} s^{-1} at leaf N concentrations of 2 g N m^{-2}. Assuming an asymptotic relationship between Amax and leaf N concentration (Evans, 1983; Sinclair & Horie, 1989), there may be scope to decrease SLN whilst maintaining Amax. Indeed, since leaves of modern wheat genotypes typically accumulate more N than 2.0 g N m^{-2} under favourable conditions (Critchley, 2001; Pask et al. 2012), NUtE could be increased by selecting for lower SLN to decrease the transient 'storage' N components of leaves. A sensitivity analysis using the wheat Sirius model predicted that decreasing SLN in the range 1-2 g m^{-2} increased NUE by 10-15% when N was limiting (Semenov et al. 2007). However under well fertilized conditions decreasing SLN below 2 g m^{-2} may not be beneficial since the SLN required for maximal RUE in field-grown winter wheat in the UK and New Zealand was estimated to be 2.1 g m^{-2} (Pask et al. 2012). Alternatively, increasing SLN above current values of 2-3 g m^{-2} seems unlikely to be advantageous overall for NUtE since leaves may operate well below light saturation in the canopy (Reynolds et al. 2000), mesophyll cell size, leaf size and light interception may be reduced (Austin et al. 1982) and many chloroplasts may end up in a light-limited state due to intra-leaf shading in thick leaves. Genetic variability in SLN is reported from 1.4-2.6 g m^{-2} for 144 durum wheat genotypes

(Araus et al. 1997), from 2.1-2.4 g m^{-2} for 17 durum wheat cultivars (Giunta et al. 2002) and from 1.4-2.2 g m^{-2} for 16 bread wheat cultivars (mean over a high and low N treatment, Moreau et al. 2012). SLN heritability in wheat is largely unknown. However, it is encouraging that the heritability for straw (leaf lamina, leaf sheath and stem) N at anthesis for winter wheat was > 0.60 under low N (Laperche et al. 2006) indicating that breeding to manipulate the amount of global canopy N should be possible.

Rubisco catalyses a wasteful reaction with oxygen that leads to the release of previously fixed CO_2 and NH_3 and the consumption of energy during photorespiration as mentioned above. Consequently, at the metabolic level, there are several avenues to increase photosynthetic efficiency. These include: (i) relaxing the photo-protected state more rapidly, (ii) reducing photorespiration through ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) with decreased oxygenase activity, (iii) the improving Rubisco activity, (iv) faster regeneration of ribulose-1,5-bisphosphate (RuBP) and (iv) introducing carbon-concentrating mechanisms associated with C4 photochemistry into C3 plants (see recent reviews by Reynolds et al. 2000; Parry et al. 2003; Long et al. 2006, Murchie et al. 2009; Zhu et al. 2010; Parry et al. 2011). These strategies all require modification of the photosynthetic components, which can only be achieved through genetic manipulation. Potential improvements in C3 cereals available from reduced photorespiration were estimated around 30 % and those from other mechanisms in the range 15-22 % (Long et al. 2006).

Alternatively, it may be possible to increase Amax by decreasing respiration in crops, although this has received less attention than photosynthesis partly due to difficulties in measurement. Respiration may consume 30% to 80% of the carbon fixed (Atkin et al. 2005) and is commonly divided into growth and maintenance components, with each exerting differing effects. Respiration, increasing with temperature and depending on phenological stage (McCullough and Hunt, 1993; Foulkes and Murchie, 2011) may be positively but non-linearly related to photosynthesis. High respiration rates (especially at night) can increase reactive oxygen species, leading to cell damage and affecting pollen viability (Prasad et al. 1999). Recent work highlighting the importance of increased night time temperature with climate change on productivity in wheat (Tester & Langridge, 2010; Lizana & Calderini, 2013) and the high sensitivity of respiration to temperature in general, suggests that the environmental responses of crop respiration to temperature is an important area on which to focus.

Post-anthesis N remobilisation and senescence dynamics

In wheat, of the N in the above-ground crop at anthesis 35-42 % is in the leaf lamina, 14-20 % in the leaf sheath, 20-31 % in the true stem and 16-23 % in the ear under optimal N supply (Pask et al. 2012; Barraclough et al. 2014; Gaju et al. 2014). Under low N conditions, the proportion of the N in the ear increases relative to that in the other plant components (Barraclough et al. 2014; Gaju et al. 2014). In field experiments in the UK and New Zealand, on winter wheat, the accumulation and remobilisation of structural, photosynthetic and

reserve N was estimated in crop components under high N and low N conditions (Pask et al. 2012). At anthesis, reserve N accounted for 44 % of above-ground N in optimally fertilised crops, and was principally located in the true stem, but was observed in all crop components at non-limiting fertiliser N treatments. The efficiency of post-anthesis N remobilisation of true stem reserve N in the true stem was low (48 %) compared to the leaf sheath (61 %) and leaf lamina (76 %), and in well fertilised crops significant quantities of non-remobilized reserve N remained in true stem at harvest.

A high capacity to absorb N in the true stem before flowering could theoretically favour a higher NUpE (Foulkes et al. 2009). In addition, favouring a greater capacity to store N in non-photosynthetic organs (*i.e.* stem internodes) may enable the translocation of a larger amount of N to grains without reducing plant photosynthetic capacity (Bertheloot et al. 2008), although the respiratory cost of maintaining a large non photosynthetic pool of storage N is unclear. In wheat, genetic variation in stem N content at anthesis is reported (Triboï and Ollier, 1991; Critchley, 2001; Pask et al. 2009; Barraclough et al. 2014; Gaju et al. 2014), as well as in post-anthesis N remobilisation efficiency from the stem (Kichey et al. 2007; Pask et al. 2009; Gaju et al. 2014). Studies in maize report early remobilisation of N from the stem before the leaf lamina (Beauchamp et al. 1976; Friedrich and Schrader, 1979). Thus high stem N remobilisation efficiency would potentially favour high NUtE through delayed senescence of the leaf lamina.

'Stay-green' phenotype refers to the capacity of a genotype to retain green leaf area for longer than a standard genotype during grain-filling (Thomas & Smart, 1993). Although under optimal conditions wheat crops are in general little limited by the assimilate supply during grain filling (Dreccer et al. 1997; Borrás et al. 2004; Calderini et al. 2006), under low to moderate N fertiliser levels there is evidence that yields can be limited by post-anthesis assimilate supply (Bogard et al. 2011; Gaju et al, 2011). Stay-green phenotypes and broader genetic variation in senescence have been reported in hexaploid wheat (Silva et al. 2000; Verma et al. 2004; Joshi et al. 2007; Christopher et al. 2008; Chen et al. 2010; Bogard et al. 2011; Chen et al. 2011; Gaju et al. 2011; Naruoka et al. 2012; Derkx et al. 2012).

Physiological mechanisms underlying these traits have not been studied extensively. Christopher et al. (2008) found that the stay-green phenotype in the spring wheat, SeriM82, was associated with extraction of deep soil water in Australia. N dynamics are an important factor in the maintenance of green leaf area in sorghum, with stay-green in sorghum hybrids linked to changes in the balance between N demand and supply during grain filling resulting in a slower rate of N translocation from the leaves to the grain (Borrell and Hammer, 2000; van Oosterom et al. 2010a, b). The latter study showed that the onset and rate of leaf senescence were explained by a supply–demand framework for N dynamics, in which individual grain N demand was sink determined and was initially met through N translocation from the stem and rachis, and then if these N pools were insufficient, from leaf N translocation. A correlation between post-anthesis N remobilisation efficiency and the onset of the rapid phase of canopy senescence was reported under low N conditions amongst 16 wheat varieties grown at

sites in the UK and France (Gaju et al. 2014). A transcription factor (*NAM-B1*) accelerates senescence and increases N remobilisation from leaves to grains in wheat (Uauy et al. 2006). Candidate regulatory genes which were members of the WRKY and NAC transcription factor families were related to senescence in controlled environment conditions (Derkx et al, 2012). In a winter wheat doubled-haploid mapping population QTLs affecting leaf senescence and grain yield and/or grain protein concentration were identified associated with QTLs for anthesis date, showing that the phenotypic correlations with leaf senescence were mainly explained by flowering time influencing post-anthesis N availability (Bogard et al. 2011).

These results suggested that a better understanding of the mechanisms determining post-anthesis N remobilisation and senescence associated with environmental characterization, particularly on their N availability during the post-anthesis period, would offer scope to raise grain yield and/or grain protein content in wheat cultivars.

Optimizing grain protein concentration and composition

Structural and metabolic proteins are present in the starchy endosperm cells of the grain, and the predominant protein fraction in this tissue is the gluten storage proteins, comprising a mixture of monomeric gliadins and polymeric glutenins. These groups of proteins are present in approximately equal amounts and together account for about 60-70 % of the total N in the endosperm tissue. The gluten proteins confer viscoelastic properties to dough crucial for processing wheat into baked food such as bread, pasta and noodles. A precise balance of gliadin and glutenin proteins is also required, as glutenins are predominantly responsible for dough elasticity (strength) required for bread-making and gliadins for dough viscosity and extensibility required for making biscuits and cakes. The qualitative composition of the grain protein is a genetic characteristic, caused in part by differences in protein synthetic capacity (Shewry and Halford, 2002; Ravel et al. 2009), whilst the rate, duration and grain protein quantitative composition (*i.e.* the ratio between the different protein fractions; Martre et al. 2003) can be modified by environmental conditions.

An inverse relationship exists between the grain protein concentration and grain yield (Kibite and Evans 1984; Simmonds, 1995, Oury et al. 2003; Oury et Godin, 2007; Bogard et al. 2010), making the simultaneous genetic improvement of yield quantity and bread-making quality a difficult task. The physiological basis of this inverse relationship relates to competition between carbon and N for energy (Munier-Jolain and Salon, 2005) and an N dilution effect by carbon based compounds (Acreche and Slafer, 2009). The grain protein deviation (GPD) is the deviation from the regression line between grain yield and grain protein concentration (GPC). GPD can be used to identify genotypes having higher GPC than expected from their GY (Monaghan et al. (2001), and it is possible to identify wheat lines that have a positive GPD amongst groups of wheat lines (Oury et al. 2003; Bogard et al. 2010; 2011). Genetic variability in GPD has been related to post-anthesis N uptake (Kichey et al. 2007; Bogard et al. 2010, 2011), and post-anthesis N uptake, in turn, is in part associated with anthesis date (Bogard et al. 2011).

Since the majority of grain N originates from remobilisation from the canopy (Pask et al. 2012; Gaju et al. 2014), rather than from post-anthesis uptake, mechanisms to enhance reserve N accumulation in the canopy and efficiency of N remobilisation should also be addressed in the genetic improvement of GPD (Hawkesford, 2014). This may be the case using the already mentioned *NAM-B1* allele (Uauy et al. 2006) that increases N remobilisation efficiency. An alternative to developing high quality and NUE wheat is to modify grain protein composition to increase dough strength and elasticity allowing for a lower GPC. In this sense, Guarda et al. (2004) observed that a decrease in GPC with year of release for cultivars introduced in Italy from 1900 to 1994 was associated with an increase in grain quality.

For wheat grown for the feed, distilling and biofuel markets (high ratio of starch to protein required), a higher NUtE will be associated with a lower GPC. The minimum GPC reported is in the range 6.8-7.2 % (Martre et al. 2006; Kindred et al. 2008; Bogard et al. 2011), equivalent (assuming a conversion ratio of 5.7 between GPC and grain N%) to 1.2-1.3 % grain N%. It is not certain whether it is possible to decrease the N % below this as for each cell in the grain there appears to be a minimum obligatory, quantitative requirement for N for the synthesis of essential amino acids and structural and metabolic proteins. This gives grain a minimum N concentration of approximately 1.5 % (Sinclair and Amir, 1992), after which, the synthesis of grain storage proteins typically increases the grain N concentration to 2.1-2.3 % (about 12-13 % protein, typical of milling wheat).

BREEDING FOR NUE

Estimation of genetic progresses

Grain yield and the N demand to maximize yield evolved simultaneously (Guarda et al. 2004; Sylvester-Bradley and Kindred 2009), leading to an equal NUE of old and recent cultivars at their respective N optimum (Sylvester-Bradley and Kindred 2009). But when old and recent varieties are compared in the same N conditions, a significant genetic improvement of NUE was measured in various studies at different N levels (Table 1).

Table 1: Assessment of yearly percent genetic gain in nitrogen use efficiency (NUE) from direct comparison of old and modern cultivars.

Period	Genotypes	N level (kg N ha^{-1})	NUE (% yr^{-1})	Reference
1962-1985	8	0	1.2	Ortiz-Monasterio et al. 1997
		75	0.4	
		150	0.6	
		300	0.9	
1977-2007	24	0	0.35	Sylvester-Bradley and Kindred 2009
		200	0.58	
1985-2010	195	150	0.37	Cormier et al. 2013
		250	0.30	

Ortiz-Monasterio et al. (1997) reported an NUE genetic progress of +0.4-1.1 % year^{-1} depending on the N levels in spring CIMMYT varieties cultivated between 1962 and 1985. Sylvester-Bradley and Kindred (2009) also reported a significant trend between +0.35-0.58 % year^{-1} comparing an old group of varieties (1977-1987) to a recent one (2001-2007) at two N levels (without N applied and with 200 kg ha-1 N applied). In the same way, Cormier et al. (2013) estimated genetic progress at +0.30-0.37 % year^{-1} between 1985 and 2010 using 195 European elite winter varieties at optimal and sub-optimal N levels. Only Muurinen et al. (2006) studying 17 spring wheat cultivar released between 1901 and 2000 observed a poorly significant genetic improvement of NUE (P = 0.055).

NUE is an integrative trait, thus its improvement could be the result of modification on several components. An increase in N harvest index (NHI) was assessed at +0.15 % year^{-1} by Brancourt-Hulmel et al. (2003) and at +0.12 % year^{-1} by Cormier et al. (2013). This improvement is independent of the semi-dwarf alleles introgressions (Gooding et al. 2012) and is associated with a decrease of N content in straw at maturity (Cormier et al. 2013). It may result from a better translocation (portion of N absorbed after anthesis and allocated to the grain) and/or a better N remobilisation. Thus, these results highlighted a breeding impact on N utilisation. An increase in N uptake was also assessed (Ortiz-Monasterio et al. 1997; Guarda et al. 2004; Sylvester-Bradley and Kindred 2009). But this conclusion has to be balanced as Foulkes et al. (1998) who studied 27 cultivars released from 1969 to 1988 and concluded that at zero N input, N offtake in grain decreased. Moreover Cormier et al. (2013) who studied a recent 214-variety panel of European elites and could not conclude on this point due to a too low genetic variance for N uptake.

To conclude, both N uptake and N utilisation may have been increased by breeding with a relative efficiency affected by the N levels (Ortiz-Monasterio et al. 1997; Le Gouis et al. 2000). We should point out that this improvement is an indirect effect of breeding for grain yield at a constant N level as no targeted selection for NUE has been conducted.

Impact of G × N interactions on direct/indirect selection efficiency

In wheat, varieties are commonly selected and registered in HN conditions. Thus, genetic progresses in LN condition results from an indirect selection. Numerous studies detected significant G × N interactions for agronomic traits (e.g. Ortiz-Monasterio et al. 1997a,b; Le Gouis et al. 2000; Laperche et al. 2006a; Barracough et al. 2010; Cormier et al. 2013) meaning that varieties genetic values differ between different N levels. Significance of G×N interactions directly affects the correlations of genetic values between different N levels, and so the best varieties at HN may not be the best at LN. Thus, when G × N interactions are significant, indirect selection efficiency (ISE) is reduced. Nevertheless, selecting at HN for LN can be efficient when heritabilities in HN are higher than in LN. Indeed, a balance between the capacity to select (heritabilities) and the genetic correlation between the environment used to select and the one where varieties will be tested is required. This balance is easy to understand when we have a look at the ISE formula (Falconer and Mackay, 1996):

$$ISE = r_{G12} \times h_2 / h_1$$

where varieties are tested in condition 1 but selected in condition 2, h_1 and h_2 are the respective heritabilities square roots in the two conditions and r_{G12} the genetic correlation between conditions, considering an equal selection intensity in both condition.

In wheat, studies reported both genetic variance decrease and environmental variance increase at LN compare to HN. Thus, heritabilities are usually lower under LN conditions (Brancourt-Hulmel et al. 2005, Laperche et al. 2006a), and indirect selection at high N can be an effective strategy to breed for low N conditions. But, few studies directly quantified this indirect selection efficiency (Brancourt-Hulmel et al. 2005; Przystalski et al. 2008; Annicchiarico et al. 2010; Cormier et al. 2013, Sarcevic et al. 2014). These studies have to be compared regarding N stresses and the number of genotypes used.

Table 2: Efficiency of selection in high N environment for low N environment (Indirect Selection Efficiency-ISE) regarding yield reduction between high and low N trials.

Genotypes	Yield reduction (%)	ISE	Reference
270	35	0.65-0.99	Brancourt-Hulmel et al. 2005
12-188	27	0.86-1.02	Przystalski et al. 2008
225	20	0.78	Cormier et al. 2013
19	10	1.04	Sarcevic et al. 2014

Using 270 breeding lines tested during two years in the same environment (northern France), Brancourt-Hulmel et al. (2005) assessed an ISE of 0.65-0.99 for grain yield with an N stress which implied a mean yield reduction of 35 % and genetic correlations between 0.83 and 0.89. Cormier et al. (2013) tested 225 commercial varieties. Comparing HN and LN, mean yield reduction was 20 % and traits heritabilities were stable. Thus, ISE was mainly dependent on genetic correlation. For grain yield it was assessed at 0.78. For the other investigated agronomic traits, ISE were between 0.25 and 0.99. The other studies used less genotypes. In Sarcevic et al. (2014), 19 varieties were tested and yield reduction was only 10 % promoting high genetic correlations. Moreover, genetic correlations were allowed to exceed 1. As results, ISE for grain yield was high (1.04) as for grain N yield (1.34) and for most of the rheological parameters (0.81-1.00) of grain quality. Analysing a dataset from seven European country comparing organic and non-organic cropping system were analyzed, Przystalski et al. (2008) assessed an ISE ranging from 0.86 to 1.02 for grain yield (calculated from the paper) under a N stress inducing a mean yield reduction of 27 %. This result seems however overestimated regarding the unbalanced dataset and the method used. Annichiario et al. (2010) studied three datasets respectively containing 7, 11, and 13 genotypes under two production systems (organic and conventional). Yield reduction ranged from 14 % to 28 % and ISE ranged from 0.89 to 1.20 for grain yield, but there were no consistent genotype by production system interactions and/or

heritabilities in organic system were lower than in conventional system mostly due to higher experimental error.

When dataset size is sufficient to properly estimate genetic correlation and an N stress is substantial, ISE for grain yield may not exceed one. Thus, regarding breeder financial issues, indirect selection is efficient in moderate N stresses however it does not overpass direct selection in LN conditions. This was already observed for maize (*Zea Mays*), for which selection under high N for performance under low N was predicted significantly less efficient than selection under low when relative yield reduction due to N stress exceeded 43 % (Bänziger et al. 1997). Concerning, varieties recommendation, the approach is different as the goal is not to increase a trait mean value but to advise wheat grower, and thus to predict which ones will be the best. In this case, we should also focus on varieties ranking between HN and LN conditions. And even when genetic correlation between HN and LN conditions are high, the probability to predict the top varieties in LN from HN ranking is low (probability of 0.55 for a genetic correlation of 0.8 in Przystalski et al. (2008) simulation study).

Molecular breeding

Molecular breeding can be defined as the use of molecular information to develop new genotypes. This molecular information can arise at different levels of the metabolic process: from gene through proteins to metabolites. In complex traits such as NUE, a lot of regulation pathways at different levels occur (*e.g.* transcription factor, post-transcriptional modification, allosteric regulation). These pathways depend on N levels (Howarth et al. 2008; Ruuska et al. 2008; Wan et al. 2013), organs (Ruuska et al. 2008), genotypes (McIntyre et al. 2011; Tenea et al. 2012), and stage (Ruuska et al. 2008; Wan et al. 2013).

In the approach to create genetically modified (GM) crop, this complexity make critical the promoter choice. Reviews of transgenic effort to improve NUE in plant were published by Pathak et al. (2011) and McAllister et al. (2012). Using the example of research on alanine aminotransferase (AlaAT), a successful transgenic approach to increase NUE in oil seed rape (Good et al. 2007) and rice (Shrawat et al. 2008) actually tested in wheat, they concluded that enzymes and proteins other than those involved in primary N uptake and assimilation may be good target potentially due to less post-transcriptional controls.

Indeed, it has been believed for a long time that due to their strategic position along the N assimilatory pathway, NR, NiR, GS, and GOGAT enzymes were major checkpoints controlling plant NUE. But, the first results of modifications of these genes had not produced completely relevant NUE phenotypes. Nevertheless, there is some evidence that increasing NR activity improves NO_2^- assimilation in Arabidopsis (Takahashi et al. 2001). Moreover, it seems that wheat genotypes exhibiting a higher NR activity have a greater potential for N utilization under non-limiting N supply with a well-coordinated system of N uptake and assimilation (Vouillot et al. 1996; Anjana et al, 2011). And recently, it was reported that overexpression of a tobacco NR gene in wheat increased the seed protein content, without the need for increased N fertilisation (Zhao et al. 2013). Such an interesting finding could rekindle the possibility of using NR

as a breeding target to improve wheat NUE, yield and grain quality. Far fewer studies have concerned the enzyme NiR in wheat.

Indirect evidence of the role of the GS enzyme in the control of NUE in wheat was also provided through correlation studies that suggested that the leaf enzyme activity could be used as a marker to monitor plant N status (Kichey et al. 2007). In addition, a number of QTLs related to grain yield and grain protein content co-localizing with structural genes encoding either cytosolic GS1 (Habash et al. 2007; Fontaine et al. 2009; Gadaleta et al. 2014) or plastidic GS2 (Gadaleta et al. 2011; Bordes et al. 2013) were identified. However, functional validation of these candidate genes will be necessary to demonstrate their impact on wheat productivity (Swarbeck et al. 2011). A recent association analysis of one of the gene encoding cytosolic GS (TaGS1a) suggest that the enzyme had an important function in the control of a number of yield-related traits (Guo et al. 2013) like its plastidic counterpart (Gadaleta et al. 2011).

Following the discovery that in rice mutants deficient in one of the two forms of NADH-GOGAT, there was a considerable reduction in spikelet number (see Yamaya and Kusano, 2014 for a review), studies on the wheat enzyme were also undertaken. Based on a quantitative genetic study in which colocalization between QTLs for NUE and the structural gene for NADH-GOGAT was observed (Quraishi et al. 2011), it was proposed that in wheat and other cereals the gene could be used to improve grain filling either using genetic manipulation, or by selecting the best alleles (Salse et al. 2013). In durum wheat, it was also found that there is a strong correlation between NADH-GOGAT gene expression and grain protein content (Nigro et al, 2013), thus indicating that unlike in a C4 plant such as maize (Martin et al. 2006), it is not cytosolic GS1 but NADH-GOGAT that is one of the major checkpoints controlling NUE in C3 cereals. Such a finding reinforces the current concept that NUE may be unique, depending not only on the species examined but also on the genetic variability within the species (Hirel et al. 2007; Simons et al. 2014).

Regarding marker assisted selection, to deal with N pathway complexity of regulation, we may think that the easiest screening would be based on protein or metabolite. Kusano et al. (2011) wrote a good review on metabolic approaches focusing on N metabolism. In wheat, only Howarth et al. (2008) assessed the impact of N supply on amino acid content during senescence. However, various proteomic studies were performed at different growing stages and organs (Bahrman et al. 2004a, 2004b, 2005; Altenbach et al. 2011; Tétard-Jones et al. 2013). But, these approaches are limited to the exploration of a narrow genetic diversity (Table 3). In fact, due to affordable cost (time and price) most of molecular information available is at the genome level as genetic molecular markers. This information was used in association mapping studies NUE related traits (Table 4) mostly using biparental design such as doubled haploids (DH) populations (An et al. 2006; Laperche et al. 2006; Habash et al. 2007; Laperche et al. 2007; Laperche et al. 2008; Fontaine et al. 2009; Li et al. 2010; Zheng et al. 2010; Bogard et al. 2011; Bogard et al. 2013) or recombinant inbred line (RIL) populations (Garcia-Suarez et al. 2010; Li et al. 2010; Guo et al. 2012; Sun et al. 2013; Xu et al. 2013).

Table 3: List of 'omics studies related to nitrogen use efficiency in wheat.

	Reference	Genotypes	N levels	Organs	Stage	Methods	data points
Proteomic	Bahrman et al. 2004a	2 (Arche, Récital)	0, 2, 8, and 20 mg N/ plant/day	leaf	60 days	2D gel electrophoresis	524 spots
	Bahrman et al. 2004b	2 (Arche, Récital)	0, 2, 8, and 20 mg N/ plant/day	leaf	60 days		541 spots
	Bahrman et al. 2005	2 (Arche, Récital)	0.5 and 3.0 mM NO_3^-	root	2nd node		860 spots
	Altenbach et al. 2011	1 (Butte 86)	0, and 30 mg N/plant/DAP	grain	maturity		54N
	Tétard-Jones et al. 2013	1 (Malacca)	organic, conventional	flag leaf	ear emergence, anthesis, kernel milk stage		111N
Transcriptomic	Ruuska et al. 2008	1 (Janz)	1 mM KNO3 and 2 mM KNO3 + 3 mM Ca(NO3)2	lower leaves and stem, flag leaf, penult internode	anthesis, 9 DPA	cDNA microarray	36,000 sequences
	Howarth et al. 2008	1 (Hereward)	48 and 192 kg N ha-1	leaf 2 and 3	senescence	GeneChip Affymetrix	55,052 transcripts
	McIntyre et al. 2011	8 (Seri × Babax pop)	0, 44, 60 and 172 kg N ha-1	stem	anthesis		
	Tenea et al. 2012	3 (Tommi, Centenaire, Cubus)	organic, conventional	flag leaf	kernel milk stage		
	Wan et al. 2013	6 (Cordiale, Hereward, Istabraq, Malacca, Marksman and Xi 19)	100, 200 and 350 kg N ha^{-1}	caryopse	14, 21, 28 and 35 DPA		
Metabolomic	Howarth et al. 2008	1 (Hereward)	48 and 192 kg N ha^{-1}	leaf 2 and 3	senescence	Gas chromatography-mass spectrometry	

Table 4: List of association mapping studies related to nitrogen use efficiency in wheat.

Reference	Pop.	Genotypes	Origin	Marker	Map (cM)	Env	Year	Site		Treatment	Traits	QTL
An et al. 2006	DH	120	Hanxuan 10 × Lumai 14	395 (AFLP, SSR, EST)	3904		2	1	2	LN=HN-150 kg N ha	5	34
Li et al. 2010	Panel +DH +RIL	260 +120 +142	Core collection Hanxuan 10 × Lumai 14 Xiaoyan 54 × Jing 411	3 TaGS2		1	1	1	2	LN HN	5	
Guo et al. 2012	RIL	131	Chuan 35050 × Shannong 483	719 (DArT, SSR, EST)	4008	12	1	1	12	N,P,K	24	380
Sun et al. 2013						3	1	1	3	NO_3^-/NH_4^+ ratio	8	147
Xu et al. 2013	RIL	182	Xiaoyan 54 × Jing 411	555 (SRR, EST, *Glu* loci)		4	2	1	2	LN HN	14	126
Laperche et al. 2007	DH	222	Arche × Recital	190 (SSR, GLU-1A/1D, Rht-B1, SPA, Fd-gogat-D1, VRN-A1, B1)	2164	14	2	4	2	LN=HN-100kg N ha		233
Laperche et al. 2006a	DH	120			2164	1	1	1			18	32
Laperche et al. 2008	DH	222			2164	14	2	4	2	LN=HN-100kg N ha	6	45
Zheng et al. 2010	DH	222		182 SSR	2164	12	2	3	2	LN HN	4	131
Fontaine et al. 2009	DH	137-221		197 (SSR)	3285	3	3	1	1		16	148
Habash et al. 2007	DH	91	CS × SQ1	449 (SSR + GS loci)	3522	1	1	1	1		21	145
Garcia-Suarez et al. 2010	RIL	114	W7984 × Opata85			4	2	1	2	LN=0 ; HN=120kg N ha	10	138
Bogard et al. 2011	DH	140	Toisondor × 3CF9107	475 (DArT, SSR, SNP)	2344	10	2	5	2	LN=(25-50)%HN	7	140
Bogard et al. 2013	3 DH	80 +80 +140	Toisondor × Quebon CF9107 × Quebon Toisondor × CF9107	741 (DArT, SSR, SNP)	2510	7	2	3	2	LN=25%HN	2	89
Bordes et al. 2013	Panel	196	Core collection	899 (DArT, SSR, SNP)		12	2	3	2	LN=HN-(35–120) kg N	8	54
Cormier et al. 2014	Panel	214	Commercial varieties	23,603 SNP	3,167	8	2	3	2	LN=HN-100 kg N	28	333

Three studies covered a broader genetic diversity (Li et al. 2010; Bordes et al. 2013; Cormier et al. 2014) using large association panels. Thus, discovering interesting quantitative trait loci these studies provided new insights on NUE genetic determinism. Indeed, QTL colocalisations with known N uptake or assimilation enzymes were assessed, but a quantity of new QTL were also discovered.

Nevertheless, several difficulties persist to implement this knowledge in breeding. Indeed, NUE and its related traits appeared highly polygenic and genetic background specific. Thus, several small loci effect should be pyramided. Moreover, information quantity will raise with the recent development of several wheat SNP arrays (90K, Wang et al. 2014; 420K, E. Paux person. comm., 670K, and 820K). Genomic prediction methods may overpass these limitations and facilitate breeding but to now these methods are still at a development stage. Added to that, G×N and more generally of G×E remain a major trade-off in marker assisted selection leading to difficulties to develop new genotypes adapted to a broad range of environments and N levels.

Prospect on new strategy: heterosis

F1 hybrid wheat cultivars have been regularly registered in Central Europe that represents more than half of the world's hybrid wheat production (Longin et al. 2012). Commercial hybrids may be produced with chemical hybridizing agents, which induce male sterility when applied at the right stage, but also based on photoperiodic sensitivity or on cytoplasmic male sterility. Limits to the use of F1 hybrids are the cost of the seed related to the difficulty to produce them on a regular basis coupled with the absence of high heterosis for yield. However hybrids may show particular characteristics for abiotic stress tolerance and NUE. Limited but consistent best-parent heterosis have been reported for grain yield under high yielding conditions, *e.g.* +4.3 % for 10 hybrids (Borghi et al. 1988), +7.3 % for 17 hybrids (Brears et al. 1988), +3.6 % for 430 hybrids (Morgan et al. 1989) in experiments conducted in field plots. On average in Europe on five studies, Longin et al. (2012) reported mid-parent heterosis around 10 %, ranging from 3.5% to 15%. It was also reported that the hybrids are more stable than pure lines (Mühleisen et al. 2014) indicating a higher tolerance to abiotic stresses.

Perezin et al. (1992) and Oury et al. (1994, 1995) reported either a higher grain protein content of the hybrids for the same yield or the same protein content despite a higher grain yield. These results tend to indicate a higher NUE and N uptake for hybrids compared to pure lines. Some studies also showed that best parent heterosis was higher at low N level than at high N level (Le Gouis and Pluchard 1996, Le Gouis et al. 2002). This was however not confirmed by Kindred and Gooding (2005) using four commercial hybrids that observed a significant heterosis only at high N level. Le Gouis et al. (2002) observed a best-parent heterosis for total N at anthesis and harvest meaning a better N uptake while Kindred and Gooding (2004) reported only little heterosis for total above-ground N but an increased N utilization efficiency. N uptake mid-parent heterosis at flowering and maturity could be related to a more efficient root system. Indeed, heterosis was shown for different root characteristics such as root length, root dry matter,

and root area (Kraljevic-Babalic et al. 1988, Wang et al. 2006, Li et al. 2013).

CONCLUSION

NUE is complex and determined by a wide diversity of physiological traits. Consequently, breeding for enhanced NUE can be achieved through selection on several components. However, compensation and regulation are numerous and dependent of the N regimes, genotypes and stage leading to difficulties to create efficient NUE phenotypes. Nevertheless, 'omics and association studies provided interesting results allowing to prioritize route of improvement. Moreover, the development of high-throughput genotyping and phenotyping methods may accelerate research on a wide diversity.

AUTHOR CONTRIBUTION

Definition of NUE and rationale for its improvement: FC and DG. Root size and morphology: JF. Root N transporter systems: BH. Interaction with micro-organisms: YML. Nitrate assimilation: BH. Leaf and canopy photosynthesis per unit N: JF. Post-anthesis N remobilisation and senescence dynamics: JF. Optimizing grain protein concentration and composition: JF. Breeding for NUE: FC and JLG. Coordination of contribution and manuscript editing: FC and JLG.

REFERENCES

Abbate, P.E., D.H. Andrade, L. Lazaro, J.H. Baraitti, H.G. Beradocco, V.H. Inza, and F. Marturano, 1998: Grain Yield in Recent Argentine Wheat cultivars. Crop Sci 38, 1203-1209.

Abenavoli, M.R., C.D. De Santis, M. Sidari, A. Sorgonà, M. Badiani, and G. Cacco, 2001: Influence of coumarin on the net nitrate uptake in durum wheat. New Phytol 150, 619-627.

Acreche, M.M., and G.A. Slafer, 2009: Variation of grain nitrogen content in relation with grain yield in old and modern Spanish wheats grown under a wide range of agronomic conditions in a mediterranean region. J. Agric. Sci. 147, 657-667.

Allard, V., P. Martre, and J. Le Gouis, 2013: Genetic variability in biomass allocation to roots in wheat is mainly related to crop tillering dynamics and nitrogen status. European Journal of Agronomy 46, 68-76.

Almario, J., Y. Moënne-Loccoz, and D. Muller, 2013: Monitoring of the relation between 2,4-diacetylphloroglucinol-producing Pseudomonas and Thielaviopsis basicola populations by real-time PCR in tobacco black root-rot suppressive and conducive soils. Soil Biol. Biochem. 57, 144-155.

Altenbach, S.B., C.K. Tanaka, W.J. Hurkman, L.C. Whiteland, W.H. Vensel, and F.M. Dupont, 2011: Differential effects of a post-anthesis fertilizer regimen on the wheat flour proteome determined by quantitative 2-DE. Proteome Science 4, 9-46.

An, D., J. Su, Q. Liu, Y. Zhu, Y. Tong, J. Li, R. Jing, B. Li, and Z. Li, 2006: Mapping QTLs for nitrogen uptake in relation to the early growth of wheat (Triticum aestivum L.). Plant Soil 284, 73-84.

Andrews, M., J.A. Raven, and P.J. Lea, 2013: Do plants need nitrate? The mechanisms by which nitrogen form affects plants. Ann Appl Biol 163, 174-199.

Angus, J.F., V.V.S.R. Gupta, G.D. Pitdson, and A.J. Good, 2014: Effects of banded ammonia and urea fertiliser on soil properties and the growth and yield of wheat. Crop Past Sci 65, 337-352.

Annicchiarico P., E. Chiapparino, and M. Perenzin, 2010: Response of common wheat varieties to organic and conventional production systems across Italian locations, and implications for selection. Field Crops Res. 116, 230-238.

Anten, N.P.R., F. Schieving, and M.J.AF. Werger, 1995: Patterns of Light and Nitrogen Distribution in Relation to Whole Canopy Carbon Gain in C3 and C4 Mono- and Dicotyledonoous Species. Oecologia 101, 504–513.

Araus, J.L., T. Amaro, Y. Zuhair, and M.M. Nachit, 1997: Effect of Leaf Structure and Water Status on Carbon Isotope Discrimination in Field Grown Durum Wheat. Plant Cell Environ 20, 1484-1494.

Atkin, O.K., D. Bruhn, V.M. Hurry, and M.G. Tjoelker, 2005: The hot and the cold: unravelling the variable response of plant respiration to temperature. Funct. Plant Biol. 32, 87-105.

Atul-Nayyar, A., C. Hamel, K. Hanson, and J. Germida, 2009: The arbuscular mycorrhizal symbiosis links

N mineralization to plant demand. Mycorrhiza 19, 239-246.

Austin, R.B., C.L. Morgan, M.A. Ford,. and S.G. Bhagwat 1982: Flag Leaf Photosynthesis of Triticum aestivum and Related Diploid and Tetraploid Species. Ann. Bot 49, 177-189.

Bahrman, N., J. Le Gouis, L. Negroni, L. Amilhat, P. Leroy, A.-L. Lainé, and O. Jaminon, 2004a: Differential protein expression assessed by two-dimensional gel electrophoresis for two wheat varieties grown at four nitrogen levels. Proteomics 4, 709–719.

Bahrman, N., L. Negroni, O. Jaminon, and J. Le Gouis, 2004b: Wheat leaf proteome analysis using sequence data of proteins separated by two-dimensional electrophoresis. Proteomics 4, 2672-2684

Bahrman, N., A. Gouy, F. Devienne-Barret, B. Hirel, F. Vedele, and J. Le Gouis, 2005: Differential change in root protein patterns of two wheat varieties under high and low nitrogen nutrition levels. Plant Science 168, 81–87.

Bai, C., Y. Liang, and M.J. Hawkesford, 2013: Identification of QTLs associated with seedling root traits and their correlation with plant height in wheat. J. Exp. Bot. 64, 1745-53.

Baldani, J. I., and V. L. D. Baldani, 2005: History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. An. Acad. Bras. Ciênc. 77, 549-579.

Bänziger, M., F.J. Betran, and H.R. Lafite 1997: Efficiency of high-nitrogen selection enviroments for improving maize for low-nitrogen target environments. Crop Science 37: 1103-1109.

Bardon, C., F. Piola, F. Bellvert, F. Z. Haichar, G. Comte, G. Meiffren, T. Pommier, S. Puijalon S., N. Tsafack, and F. Poly, 2014: Evidence for biological denitrification inhibition (BDI) by plant secondary metabolites. New Phytol. DOI: 10.1111/nph.12944.

Barraclough, P.B., H. Kuhlmann, and A.H. Weir, 1989: The effects of prolonged drought and nitrogen fertiliser on root and shoot growth and water uptake by winter wheat. J. Agron. Crop Sci. 163, 352-360.

Barraclough, P.B., J.R. Howarth, J. Jones, R. Lopez-Bellido, S. Parmar, C.E. Shepherd, and M.J. Hawkesford , 2010: Nitrogen efficiency of wheat: genotypic and environmental variation and prospects for improvement. Eur J Agron 33, 1-11.

Barraclough, P.B., R. Lopez-Bellido, and M.J. Hawkesford, 2014: Genotypic variation in the uptake, partitioning and remobilisation of nitrogen during grain-filling in wheat. Field Crops Research 156, 242-248.

Beauchamp, E.G., L.W. Kannenberg, and R.B. Hunter, 1976: Nitrogen accumulation and

translocation in corn genotypes following silking. Agron. J. 68, 418–422.

Behl, R. K., S. Ruppel, E. Kothe, and N. Narula, 2012: Wheat x Azotobacter x VA Mycorrhiza interactions towards plant nutrition and growth – a review. J. Appl. Bot. Food Qual. 81, 95-109.

Berendse, F., and R. Aerts, 1987: Nitrogen use efficiency: a biologically meaningful definition? Functional Ecology 1, 293-296.

Bernard, S.M., A.L. Møller, G. Dionisio, T. Kichey, T.P. Jahn, F. Dubois, M. Baudo, M.S. Lopes, T. Tercé-Laforgue, C.H. Foyer, M.A. Parry, B.G. Forde, J.L. Araus, B. Hirel, J.K. Schjoerring, and D.Z. Habash, 2008: Gene expression, cellular localisation and function of glutamine synthetase isozymes in wheat (*Triticum aestivum* L.). Plant Mol Biol 67, 89-105.

Bertheloot, J., P. Martre, B. Andrieu, 2008: Dynamics 1 of light and nitrogen distribution during grain filling within wheat canopy. Plant Physiol 148, 1707-1720.

Bertheloot, J., P.H. Cournède, and B. Andrieu, 2011: NEMA, a functional–structural model of nitrogen economy within wheat culms after flowering. I. Model description. Ann Bot 108, 1085-1096.

Bertrand, H., C. Plassard, X. Pinochet, B. Touraine, P. Normand, and J. C. Cleyet-Marel, 2000: Stimulation of the ionic transport system in Brassica napus by a plant growth-promoting rhizobacterium (Achromobacter sp.). Can. J. Microbiol. 46, 229-236.

Bingham, I., A. Karley, P. White, W. Thomas, and J. Russell, 2012: Analysis of improvements in nitrogen use efficiency associated with 75 years of spring barley breeding. Eur J Agron 42, 49-58

Bindraban, P.S., 1999: Impact of canopy nitrogen profile in wheat on growth. Field Crops Res 63, 63-77.

Bloom, A.J., M. Burger, B.A. Kimball, and Jr.P. Pinter, 2014: Nitrate assimilation is inhibited by elevated CO2 in field grown wheat. Nature Clim Chang 4, 477-480.

Blum, A., 2009: Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. Field Crops Research 112, 119–123.

Bogard, M., V. Allard, M. Brancourt-Hulmel, E. Heumez, J.M. Machet, M.H. Jeuffroy, P. Gate, P. Martre , and J. Le Gouis, 2010: Deviation from

the grain protein concentration–grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. Journal of experimental botany 61, 4303-4312.

Bogard, M., M. Jourdan, V. Allard, P. Martre, M.R. Perretant, C. Ravel, E. Heumez, S. Orford, J. Snape, S. Griffiths, O. Gaju, J. Foulkes, and J. Le Gouis, 2011: Anthesis date mainly explained correlations between post-anthesis leaf senescence, grain yield, and grain protein concentration in a winter wheat population segregating for flowering time QTLs. J. Exp. Bot.62, 3621–3636.

Bogard, M., V. Allard, P. Martre, E. Heumez, J.W. Snape, S. Griffiths, O. Gaju, J. Foulkes, and J. Le Gouis, 2013: Identifying wheat genomic regions for improving grain protein concentration independently of grain yield using multiple inter-related populations. Molecular Breeding 31, 587-599.

Boisson, M., K. Mondon, V. Torney, N. Nicot, A.L. Laine, N. Bahrman, A. Gouy, F. Daniel-Vedele, B. Hirel, P. Sourdille, M. Dardevet, C. Ravel, J. Le Gouis, 2005: Partial sequences of nitrogen metabolism genes in hexaploid wheat. Theor Appl Genet 110, 932-940.

Bonkowski, M., 2004: Protozoa and plant growth: the microbial loop in soil revisited. New Phytol. 162, 617-631.

Bordes, J., C. Ravel, J.P. Jaubertie, B. Duperrier, O. Gardet, E. Heumez, A.L. Pissavy, G. Charmet, J. Le Gouis, and F. Balfourrier, 2013: Genomic regions associated with the nitrogen limitation response revealed in a global wheat core collection. Theor Appl Genet 126, 805-822

Borghi, B., M. Perenzin and R. J. Nash 1988. Agronomic and qualitative characteristics of ten bread wheat hybrids produced using a chemical hybridizing agent. Euphytica 39:185-194.

Borras, L., Slafer, G.A., and M.E. Otegui, 2004: Seed dry weight response to source–sink

manipulations in wheat, maize and soybean: a quantitative reappraisal. Field Crop Res. 86, 131-146.

Borrell, A.K., and G.L. Hammer, 2000: Nitrogen dynamics and the physiological basis of

stay-green in sorghum. Crop Sci. 40, 1295–1307.

Bottini, R., F. Cassán, and P. Piccoli, 2004: Gibberellin production by bacteria and its involvement in plant growth promotion and yield increase. Appl. Microbiol. Biotechnol. 65, 497-503.

Bradley, P.M., and J.T. Morris, 1991: The influence of salinity on the kinetics of uptake in Spartina alterniflora. Oecologia 85, 375-380.

Brancourt-Hulmel, M., G. Doussinault, C. Lecomte, P. Bérard, B. LeBuanec, and M. Trottet, 2003: Genetic improvement of agronomic traits of winter wheat cultivars released in France from 1946 to 1992. Crop Science 43: 37- 45.

Brancourt-Hulmel, M., E. Heumez, P. Pluchard, D. Beghin, C. Depatureaux, A. Giraud, J. Le Gouis, 2005: Indirect versus direct selection of winter wheat for low input or high input levels. Crop Sci 45, 1427-1431.

Brazelton, J. N., E. E. Pfeufer, T. A. Sweat, B. B. McSpadden Gardener, and C. Coenen, 2008: 2,4-diacetylphloroglucinol alters plant root development. Mol. Plant-Microbe Interact. 21, 1349-1358.

Brears, T., A.G. Hydon and, J. Bingham 1988: An assessment of the feasability of producing F1 and F2 hybrids for the UK. Proc 7th Int Wheat Genet Symp: 1057-1062.

Buchner, P., and M.J. Hawkesford, 2014: Complex phylogeny and gene expression patterns of members of the NITRATE TRANSPORTER1/PEPTIDE TRANSPORTER family (NPF) in wheat. J Exp Bot 65, 5697-57101.

Buée, M., W. De Boer, F. Martin, L. van Overbeek, and E. Jurkevitch, 2009: The rhizosphere zoo: an overview of plant-associated communities of microorganisms, including phages, bacteria, archaea, and fungi, and of some of their structuring factors. Plant Soil 321, 189-212.

Cacco, G., E. Attinà, A. Gelsomino, and M. Sidari, 2000: Effect of nitrate and humic substances of different molecular size on kinetic parameters of nitrate uptake in wheat seedlings. J Plant Nutr Soil Sci 163, 313-320.

Cai, C., X.Q. Zhao, Y.G. Zhu, B. Li, Y.P. Tong, and Z.S. Li, 2007: Regulation of the high-affinity nitrate transport system in wheat roots by exogenous abscisic acid and glutamine. J Integr Biol 49, 1719-1725.

Calderini, D.F., M.P. Reynolds, and G.A. Slafer, 2006: Source–sink effects on grain weight

of bread wheat, durum wheat, and triticale at different locations. Aust. J. Agric.

Res. 57, 227–233.

Carvalho, P., and M.J. Foulkes, 2011: Roots and the uptake of water and nutrients. In: Meyers RA, ed. Encyclopedia of Sustainability Science and

Technology. Springer, Heidelberg, Germany, Chapter 195, 1390-1404.

Carvalho, P., S. Azam-Ali, and M.J. Foulkes, 2014: Quantifying relationships between rooting traits and water uptake under drought in Mediterranean barley and durum wheat. Journal of Integrative Plant Biology 56, 455-469.

Cassán, F., D. Perrig, V. Sgroy, O. Masciarelli, C. Penna, and V. Luna, 2009: Azospirillum brasilense Az39 and Bradyrhizobium japonicum E109, inoculated singly or in combination, promote seed germination and early seedling growth in corn (Zea mays L.) and soybean (Glycine max L.). Eur. J. Soil Biol. 45, 28-35.

Causin, H.F., and A.J. Barneix, 1993: Regulation of NH4+, uptake in wheat plant: effect of root ammonium concentration and amino acids. Plant Soil 151, 211-218.

Chen, J.B., Y. Liang, X.Y. Hu, X.X. Wang, F.Q. Tan, H.Q. Zhang, Z.L. Ren, and P.G. Luo, 2010: Physiological characterization of 'stay green' wheat cultivars during the grain filling stage under field growing conditions. Acta Physiol Plant 32, 875-882.

Chen, C.C., G.Q. Han, H.Q. He, and M. Westcott, 2011: Yield, protein, and remobilization of water soluble carbohydrate and nitrogen of three spring wheat cultivars as influenced by nitrogen input. Agron. J.103, 786-795.

Cheng, W., D. W. Johnson, and S. Fu, 2003: Rhizosphere effects on decomposition. Soil Sci. Soc. Am. J. 67, 1418-1427.

Christiansen, M.W., and P.L. Gregersen, 2014: Members of the barley NAC transcrition factor gene family show differential co-regulation with senescence-associated genes during senescence of flag leaves. J Exp Bot 65, 4009-4022.

Christiansen-Weniger, C., A. F. Groneman, and J. A. van Veen, 1992: Associative N2 fixation and root exudation of organic acids from wheat cultivars of different aluminium tolerance. Plant Soil 139, 167-174.

Christopher, J.T., A.M. Manschadi, G.L. Hammer, and A.K. Borell, 2008: Developmental and physiological traits associated high yield and stay-green phenotype in wheat. Aust. J. Agric. Res. 59, 354-364.

Coelho, M. R. R., I. E. Marriel, S. N. Jenkins, C. V. Lanyon, L. Seldin, and A. G. O'Donnell, 2009: Molecular detection and quantification of nifH gene sequences in the rhizosphere of sorghum (Sorghum bicolor) sown with two levels of nitrogen fertilizer. Appl. Soil Ecol. 42, 48-53.

Cohan, JP., 2009 : Prix des engrais azotés : Quels impacts sur les céréales ? Perspectives Agricoles 352, 18-22.

Cohen, A. C., R. Bottini, and P. N. Piccoli, 2008: Azospirillum brasilense Sp 245 produces ABA in chemically-defined culture medium and increases ABA content in arabidopsis plants. Plant Growth Regul. 54, 97-103.

Combes-Meynet, E., J. F. Pothier, Y. Moënne-Loccoz, and C. Prigent-Combaret, 2011: The Pseudomonas secondary metabolite 2,4-diacetylphloroglucinol is a signal inducing rhizoplane expression of Azospirillum genes involved in plant-growth promotion. Mol. Plant-Microbe Interact. 24, 271-284.

Cormier, F., S. Faure, P. Dubreuil, E. Heumez, K. Beauchêne, S. Lafarge, S. Praud, and J. Le Gouis, 2013: A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (*Triticum eastivum* L.). Theor Appl Genet 126, 3035–3048.

Cormier, F., J. Le Gouis, P. Dubreuil, S. Lafarge, and S. Praud, 2014: A genome-wide identification of chromosomal regions determining nitrogen use efficiency components in wheat (*Triticum eastivum* L.). Theor Appl Genet DOI:10.1007/s00122-014-2407-7.

Couillerot, O., E. Combes-Meynet, J. F. Pothier, F. Bellvert, E. Challita, M. A. Poirier, R. Rohr, G. Comte, Y. Moënne-Loccoz, and C. Prigent-Combaret, 2011: The role of the antimicrobial compound 2,4-diacetylphloroglucinol in the impact of biocontrol Pseudomonas fluorescens F113 on Azospirillum brasilense phytostimulators. Microbiology 157, 1694-1705.

Crawford, N.M., and A.D.M. Glass, 1998: Molecular and physiological aspects of nitrate uptake in plants. Trends in Plant Sci 3, 389-395.

Creus, C. M., M. Graziano, E. M., Casanovas, M. A. Pereyra, M. Simontacchi, S. Puntarulo, C. A. Barassi, and L. Lamattina, 2005: Nitric oxide is involved in the Azospirillum brasilense-induced lateral root formation in tomato. Planta 221, 297-303.

Criddle, R.S., M.R. Ward, and R.C. Huffaker, 1988: Nitrogen uptake by wheat seedlings, interactive effects of four nitrogen sources: NO3-, NO2-, NH4+, and urea. Plant Physiol 86, 166-175.

Critchley, C.S., 2001: A Physiological Explanation for the Canopy Nitrogen Requirement of Wheat. Ph.D. Thesis, University of Nottingham, UK.

Curzi, M. J., C. M. Ribaudo, G. D. Trinchero, J. A. Cura, and E. A. Pagano, 2008: Changes in the content

of organic and amino acids and ethylene production of rice plants in response to the inoculation with Herbaspirillum seropedicae. J. Plant Interac. 3, 163-173.

Dechorgant, J., C.T. Nguyen, P. Armengaud, M.J. Jossier, E. Diatloff, S. Filleur, and F. Daniel-Vedele, 2011: From the soil to the seeds: along journey of nitrate in plants. J Exp Bot 62, 1349-1359.

Deckard EL, Bush RH. 1978. Nitrate reductase assays as prediction test for crosses and lines in spring wheat. Crop Sci. 18, 289-293.

Den Herder, G., Van Isterdael, G., Beeckman, T., and De Smet, I., 2010: The roots of a new green revolution. Trend in Plant Science 15:600-607.

Derkx, A.P., S. Orford, S. Griffiths, M.J. Foulkes, and M.J. Hawkesford, 2012: Identification of differentially senescing mutants of wheat and impacts on yield, biomass and nitrogen partitioning. J. Integr. Plant Biol. 54, 555–566.

Dobbelaere, S., A. Croonenborghs, A. Thys, A. Vande Broek, and J. Vanderleyden, 1999: Phytostimulatory effect of Azospirillum brasilense wild type and mutant strains altered in IAA production on wheat. Plant Soil 212, 153-162.

Dreccer, M., A. Schapendonk, G. Slafer, and R. Rabbinge, 2000: Comparative response of wheat and oilseed rape to nitrogen supply: absorption and utilisation efficiency of radiation and nitrogen during the reproductive stages determining yield. Plant Soil 220, 189-205.

Dubois, F., T. Tercé-Laforgue, M.B. Gonzalez-Moro, M.B. Estavillo, R. Sangwan, A. Gallais, and B. Hirel, 2003: Glutamate dehydrogenase in plants: is there a new story for an old enzyme? Plant Physiol Biochem 41, 565-576.

Ehdaie, B, A.E. Hall, G.D. Farquhar, H.T. Nguyen, and J.G. Waines, 1991: Water-use efficiency and carbon isotope discrimination in wheat. Crop Sci. 31: 1282–1288.

Ehdaie B, and J.G. Waines, 1993: Variation in water-use efficiency and its components in wheat. I. Well-watered pot experiment. Crop Sci. 33, 294–299.

Ehdaie, B., 1995: Variation in water-use efficiency and its components inwheat. II. Pot and field experiments. Crop Sci. 35, 1617-26.

Ehdaie, B., and J.G. Waines, 1997: Growth and evapotranspiration efficiency in landrace and dwarf spring wheats. J. Genet. Breeding 51, 201-209.

Evans, J.R., 1983: Nitrogen and photosynthesis in the flag leaf of wheat (Triticum aestivum L.). Plant Physiol 72, 297-302.

Fageria, N.K., V.C. Baligar, and Y.C. Li 2008: The role of nutrient efficient plants in improving crop yields in the twenty first century. J. Plant Nutr. 31, 1121-1157.

Falconer, D., and T. Mackay, 1996: Introduction to quantitative genetics. 4th ed. Longman Scientific & Technical, New York.

Feil, B., 1997. The inverse yield-protein relationships in cereals: possibilities and limitations for genetically improving the grain protein yield. Trends in Agronomy 1: 103-119.

Fester, T., W. Maier, and D. Strack, 1999: Accumulation of secondary compounds in barley and wheat roots in response to inoculation with an arbuscular mycorrhizal fungus and co-inoculation with rhizosphere bacteria. Mycorrhiza 8, 241-246.

Field, C., 1983: Allocating Leaf Nitrogen for the Maximzation of Carbon Gain: Leaf Age as a Control on the Allocation Program. Oecologia 56, 341–347.

Fontaine, J.X., C. Ravel, K. Pageau, E. Heumez, F. Dubois, B. Hirel, and J. Le Gouis, 2009: A quantitative genetic study for elucidating the contribution of glutamine synthetase, glutamate dehydrogenase and other nitrogen-related physiological traits to the agronomic performance of common wheat. Theor Appl Genet 119, 645-662.

Fontaine, J.X., T. Tercé-Laforgue, P. Armengaud, G. Clément, J.P. Renou, S. Pelletier, M. Catterou, M. Azzopardi, Y. Gibon, P.J. Lea PJ, B. Hirel, and F. Dubois, 2012: Characterization of a NADH-dependent glutamate dehydrogenase mutant of Arabidopsis demonstrates the key role of this enzyme in root carbon and nitrogen metabolism. The Plant Cell 24, 4044-4065.

Ford, K.E., P.J. Gregory, M.J. Gooding, and S. Pepler, 2006: Genotype and fungicide effects on late-season root growth of winter wheat. Plant Soil 284, 33-44.

Forde, B.J., and P.J. Lea, 2007: Glutamate in plants: metabolism, regulation, and signalling. J Exp Bot 58, 2339-2358.

Foulkes, M.J., R. Sylvester-Bradley, and R.K. Scott, 1998. Evidence for differences between winter wheat cultivars in acquisition of soil mineral nitrogen and uptake and utilization of applied fertiliser nitrogen. J. Agric. Sci 130, 29–44.

Foulkes, M.J., M.J. Hawkesford, P.B. Barraclough, M.J. Holdsworth, S. Kerr, S. Kightley, and P.R.

Shewry, 2009: Identifying traits to improve the nitrogen economy of wheat: Recent advances and future prospects. Field Crops Research 114, 329-342.

Foulkes, M.J., and E.H. Murchie, 2011: Optimising canopy physiology traits to improve the nutrient-utilization efficiency of crops. In: The molecular basis of nutrient use efficiency in crops. Ed M. Hawkesfrod and P. Barraclough, pp 65-82, Wiley - Blackwell.

Friedrich, J.W., and L.E. Schrader, 1979: N deprivation in maize during grain-filling. II.

Remobilisation of 15N and 35S and the relationship between N and S accumulation. Agron. J. 71, 466–472.

Gaju, O., V. Allard, P. Martre, J.W. Snape, E. Heumez, J. Le Gouis, D. Moreau, M. Bogard, S. Griffiths, S. Orford, S. Hubbart, and M.J. Foulkes, 2011: Identification of traits to improve the nitrogen-use efficiency of wheat genotypes. Field Crops Res. 123, 139-152.

Gaju, O., V. Allard, P. Martre, J. Le Gouis, D. Moreau, M. Bogard, S. Hubbart, and M.J. Foulkes, 2014: Nitrogen partitioning and remobilization in relation to leaf senescence, grain yield and grain N concentration in wheat cultivars. Field Crops Res 155, 213-223.

Garcia-Suarez, J.V., M.S. Roder, and J.L. Diaz de León, 2010: Identification of QTLs and associated molecular markers of agronomic traits in wheat (Triticum aestivum L.) under two conditions of nitrogen fertilization. Cereal Research Communications 38, 459-470.

Garnett, T., V. Conn, and B. Kaiser, 2009: Root based approach to improving nitrogen use efficiency in plants. Plant Cell Envir 32, 1272-1283.

Gaur, V.S., U.S. Singh, A.K. Gupta, and A. Kumar, 2012: Understanding the differential nitrogen sensing mechanism in rice genotypes through the expression analysis of high and low affinity ammonium transporter genes. Mol Biol Rep 39, 2233-2241.

Germida, J., and S. Siciliano, 2001: Taxonomic diversity of bacteria associated with the roots of modern, recent and ancient wheat cultivars. Biol. Fertil. Soils 33, 410-415.

Glass, A.D.M., 2009: Nitrate uptake by plant roots. Botany 87, 659-667.

Gioseffi, E., A. de Neergaard, and J.K. Schjoerring, 2012: Interactions between uptake of amino acids and inorganic nitrogen in wheat plants. Biogeosciences 9, 1509–1518.

Girin, T., L.C. David, C. Chardin, R. Sibout, A. Krapp, S. Ferrario-Mery, F. Daniel-Vedele, 2014: Brachypodium: a promising hub between model species and cereals. J Exp Bot 65, 5683-5696.

Giunta, F., M. Rosella, and M. Deidda, 2002: SPAD readings and associated leaf traits in durum wheat, barley and triticale cultivars. Euphytica 125, 197-205.

Glick, B. R., 2005: Modulation of plant ethylene levels by the bacterial enzyme ACC deaminase. FEMS Microbiol. Lett. 251, 1-7.

Good A.G., A.K. Shrawat, and D.G. Muench, 2004: Can less yield more? Is reducing nutrient input into the environment compatible with maintaining crop production? Trends in Plant Science 9, 597–605.

Good, A.G., S.J. Johnson, M. De Pauw, R.T. Carroll, N. Savidov, J. Vidmar, Z. Lu, G. Taylor, and V. Stroeher, 2007: Engineering nitrogen use efficiency with alanine aminotransferase. Can. J. Bot. 85, 252–262.

Gooding, M.J., W.P. Davies, 1992: Foliar urea fertilization of cereals: a review. Fertil Res 32, 209-222.

Gooding, M. J., M. Addisu, R.K. Uppal, J.W. Snape, and H.E. Jones, 2012: Effect of wheat dwarfing genes on nitrogen-use efficiency. J Agric Sci 150, 3-22.

Goyal, S.S., and R.C. Huffaker, 1986: The uptake of NO_3^-, NO_2^-, and NH_4^+ by intact wheat (*Triticum aestivum*) seedlings. Plant Physiol 82, 1051-1056.

Graham, J. H., and R. G. Linderman, 1980: Ethylene production by ectomycorrhizal fungi, Fusarium oxysporum f. sp. pini, and by aseptically synthesized ectomycorrhizae and Fusarium-infected Douglas-fir roots. Can. J. Microbiol. 26, 1340-1347.

Gregory, P.J., and S.C. Brown, 1989: Root growth, water use and yield of crops in dry environments: what characters are desirable? Aspects Appl. Biol. 22, 234-243.

Gu, R., F. Duan, X. An, F. Zhang, N. von Wirèn, and L. Yuan, 2013: Characterization of the AMT-mediated high affinity ammonium uptake in roots of maize (Zea mays L.). Plant Cell Physiol 54, 1515-1524.

Guarda, G., S. Padovan, and G. Delogu, 2004: Grain yield, nitrogen-use efficiency and baking quality of old and modern Italian bread-wheat cultivars grown at different nitrogen levels. Eur J Agron 21,181-192.

Guo,Y., F.M. Kong, Y.F. Xu,Y. Zhao, X. Liang, Y.Y. Wang, D.G. An, and S.S. Li, 2012: QTL mapping for seedling traits in wheat grown under varying

concentrations of N, P and K nutrients. Theoretical applied genetics 124, 851-865.

Habash, D.Z., S. Bernard, J. Schondelmaier, J. Weyen, and S.A. Quarrie, 2007: The genetics of nitrogen use in hexaploid wheat : N utilisation, development and yield. Theor Appl Genet 114, 403-419.

Hakwesford, M.J., 2014: Reducing the reliance on nitrogen fertilizer for wheat production J Cereal Sci. 59, 276-283.

Hall, A.J., and R.A. Richards, 2013: Prognosis for genetic improvement of yield potential and water-limited yield of major grain crops Field Crops Res. 143, 18-33.

Hamada, A., M. Nitta, S. Nasuda, K. Kato, M. Fujita, H. Matsunaka, and Y. Okumoto, 2012: Novel QTLs for growth angle of seminal roots in wheat (Triticum aestivum L.). Plant Soil 354, 395-405.

Hawkins, H.G., E. George, 2001: Reduces 15N-nitrogen transport through arbuscular hyphae to Triticum aestivum supplied with ammonium vs. nitrate nutrition. Ann Bot 87, 303-311.

Herold, M. B., E. M. Baggs, and T. J. Daniell, 2012: Fungal and bacterial denitrification are differently affected by long-term pH amendment and cultivation of arable soil. Soil Biol. Biochem. 54, 25-35.

Hirose, T., 1971: Nitrogen turnover and dry matter production of a *Solidago altissima* population. Jpn J Ecol 21, 18–32

Hirose, T., and M.J.A. Werger, 1987: Maximizing daily canopy photosynthesis with respect to the leaf nitrogen allocation pattern in the canopy. Oceologia 72, 520-526.

Hirose, T., 2011: Nitrogen use efficiency revisited. Oecologia 166, 863-867.

Hoad, S.P., G. Russell, M.E. Lucas, and I.J. Bingham, 2001: The management of wheat, barley and oat root systems. In: Sparks DL, ed. Advances in Agronomy. Elsevier Academic Press Inc, San Diego. pp. 193-246.

Hochholdinger, F., and R. Tuberosa, 2009: Genetic and genomic dissection of maize root development and architecture. Curr. Opin. Plant Biol. 12, 172-177.

Hooda, P.S., A.C. Edwards, H.A. Anderson, and A. Miller, 2000: A review of water quality concerns in livestock farming areas. Sci Tot Environ 250, 143-167.

Howarth, J.R., S. Parmar, J. Jones, C. Shepherd, D.I. Corol, A.M. Galster, N.D. Hawkins, S.J. Miller, J.M. Baker, P.J. Verrier, J.L. Ward, M.H. Beale, P.B. Barraclough, and M.J. Hawkesford, 2008: Co-ordinated expression of amino acid metabolism in response to N and S deficiency during wheat grain filling. J Exp Bot 59, 3675-3689.

Hungria, M., R. J. Campo, E. M. Souza, and F. O. Pedrosa, 2010: Inoculation with selected strains of Azospirillum brasilense and A. lipoferum improves yields of maize and wheat in Brazil. Plant Soil 331, 413-425.

Hurd, E. 1964: Root study of three wheat varieties and their resistance to drought and damage by soil cracking. Can. J. Plant Sci. 44, 240-248.

Hsu, S. -F., and D. H. Buckley, 2009: Evidence for the functional significance of diazotroph community structure in soil. ISME J. 3, 124-136.

Iniguez, A. L., Y. Dong, and E. W. Triplett, 2004: Nitrogen fixation in wheat provided by Klebsiella pneumoniae 342. Mol. Plant-Microbe Interact. 17, 1078-1085.

Jarvis, S., N. Hutchings, F. Brentrup, J.E. Olesen, and K.W.van de Hoek, 2011: Nitrogen flows in framing systems across Europe. In: B. Grizzetti (ed.), The European Nitrogen Assessment. Sources, Effects and Policy Perspectives, 21-28. Cambridge University Press.

Jeuffroy, M. H., B. Ney, and A. Ourry, 2002: Integrated physiological and agronomic modelling of N capture and use within the plant. Journal of Experimental Botany 53, 809-823.

Joshi AK, M. Kumari, V.P. Singh, C.M. Reddy, S. Kumar, J. Rane, and R. Chand, 2007: Stay green trait: Variation, inheritance and its association with spot blotch resistance in spring wheat (Triticum aestivum L.). Euphytica 153, 59–71.

Kaiser, W.M., E. Planchet, S. Rümer, 2011: Nitrate reductase and nitric oxide. In: C.H. Foyer, H. Zhang (ed.), Annual Plant Reviews, Nitrogen Metabolism in Plants in the Post-genomic Era, 127-146. Wiley-Blackwell: Chichester.

Kapulnik, Y., Y. Okon, and Y. Henis, 1987: Yield response of spring wheat cultivars (Triticum aestivum and T. turgidum) to inoculation with Azospirillum brasilense under field conditions. Biol. Fertil. Soils 4, 27-35.

Karamos, R.E., K. Hanson, and F.C. Stevenson, 2014: Nitrogen form, time and rate of application, and nitrification inhibitor effects on crop production. Can J Plant Sci 94, 425-432.

Kibite, S., and L.E. Evans, 1984: Causes of negative correlations between grain yield and grain protein concentration in common wheat. Euphytica 33, 801–810.

Kichey, T., J. Le Gouis, B. Hirel, and F. Dubois, 2005: Evolution of the cellular and subcellular localization of glutamine synthetase and glutamate dehydrogenase during flag leaf senescence in wheat (*Triticum aestivum* L.). Plant Cell Physiol 46, 964-974.

Kichey, T., B. Hirel, E. Heumez, F. Dubois, and J. Le Gouis, 2007: In winter wheat (Triticum aestivum L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. Field Crops Res. 102, 22-32.

Kindred, D.R. and M.J. Gooding, 2004: Heterotic and seed rate effects on nitrogen efficiencies in wheat. Journal of Agricultural Science, Cambridge 142:639-657.

Kindred, D.R. and M.J. Gooding, 2005: Heterosis for yield and its physiological determinants in wheat. Euphytica 142, 149-159.

Kindred, D.R., T.M.O. Verhoeven, R.M. Weightman, J.S. Swanston, R.C. Agu, J.M. Brosnan, and R. Sylvester-Bradley, 2008: Effects of variety and fertiliser nitrogen on alcohol yield, grain yield, starch and protein content, and protein composition of winter wheat Journal of Cereal Science 48, 46–57.

Kong, L., F. Wang, L. López-bellido, J.M. Garcia-Mina, and J. Si, 2013: Agronomic improvement through the genetic and physiological regulation of nitrogen uptake in wheat (*Triticum aestivum* L.). Plant Biotech Rep 7, 129-139.

Körschens, M., E. Albert, M. Armbruster, D. Barkusky, M. Baumecker, L. Behle-Schalk, R. Bischoff, Z. Čergan, F. Ellmer, F. Herbst, S. Hoffmann, B. Hofmann, T. Kismanyoky, J. Kubat, E. Kunzova, C. Lopez-Fando, I. Merbach, W. Merbach, M.T. Pardor, J. Rogasik, J. Rühlmann, H. Spiegel, E. Schulz, A. Tajnsek, Z. Toth, H. Wegener, and W. Zorn, 2013: Effect of mineral and organic fertilization on crop yield, nitrogen uptake, carbon and nitrogen balances, as well as soil organic carbon content and dynamics: results from 20 European long-term field experiments of the twenty-first century. Arch Agron Soil Sci 59, 1017-1040.

Kraljevic-Balalic, M., R. Kastori, and M. Vojvodic, 1988: Inheritance of total root area, length and dry weight in F1 wheat crosses. Genetika 20, 229-234.

Kull, O. 2002: Acclimation of photosynthesis in canopies: models and limitations. Oecologia 133, 267-279.

Labboun, S., T. Tercé-Laforgue, A. Roscher, M. Bedu, F.M. Restivo, C.N. Velanis, D.S. Skopelitis, P.N. Moshou, K.A. Roubelakis-Angelakis, A. Suzuki, and B. Hirel, 2009: Resolving the role of plant glutamate dehydrogenase: I. In vivo real time nuclear magnetic resonance spectroscopy experiments. Plant Cell Physiol 50, 1761-1773.

Lammerts van Bueren, E. T., S. S. Jones, L. Tamm, K. M. Murphy, J. R. Myers, C. Leifert, and M. M. Messmer, 2011: The need to breed crop varieties suitable for organic farming, using wheat, tomato and broccoli as examples: A review. NJAS-Wagen. J. Life Sci. 58, 193-205.

Laperche, A., F. Devienne-Barret, O. Maury, J. Le Gouis, and B. Ney, 2006a: A simplified conceptual model of carbon/nitrogen functioning for QTL analysis of wheat adaptation to nitrogen deficiency. Theor Appl Genet 113, 1131-1146.

Laperche, A., M. Brancourt-Hulmel, E. Heumez, O. Gardet, and J. Le Gouis, 2006b: Estimation of genetic parameters of a DH wheat population grown at different N stress levels characterized by probe genotypes. Theoret Appl Genet 112, 797-807.

Laperche, A., M. Brancourt-Hulmel, E. Heumez, O. Gardet, E. Hanocq, F. Devienne-Barret, and J. Le Gouis, 2007: Using genotype x nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints. Theor Appl Genet 115, 399–415

Laperche, A., J. Le Gouis, E. Hanocq, and M. Brancourt-Hulmel, 2008: Modelling nitrogen stress with probe genotypes to assess genetic parameters and genetic determinism of winter wheat tolerance to nitrogen constraint. Euphytica 161, 259–271.

Le Gouis, J. , and P. Pluchard, 1996: Genetic variation for nitrogen use efficiency in winter wheat (Triticum aestivum L.). Euphytica 92, 221-224.

Le Gouis, J., B. Beghin, E. Heumez, and P. Pluchard, 2000: Genetic differences for nitrogen uptake and nitrogen utilisation efficiencies in winter wheat. Eur J Agron 12, 163-173.

Le Gouis, J., D. Béghin, E. Heumez, and P. Pluchard, 2002: Diallel analysis of winter wheat at two nitrogen levels. Crop Science 42, 1129-1134.

Lea, P.J., and R.A. Azevedo, 2007: Nitrogen use efficiency. 2. Amino acid metabolism. Ann Appl Biol 151, 269-275.

Lea, P.J., and B.J. Miflin, 2011: Nitrogen assimilation and its relevance to crop improvement. In: H. Zhang (ed.), Annual Plant Reviews, Nitrogen Metabolism in Plants in the Post-genomic Era, 1-40. Wiley-Blackwell, Chichester.

Lemaire, G., E. Oosterom, J. Van Sheehy, M.H. Jeuffroy, A Massignam, and L. Rossat, 2007: Is crop N demand more closely related to dry matter accumulation or leaf area expansion during vegetative growth ? Field Crops Research 100, 91–106.

Léran, S., K. Varala, J.C. Boyer, M. Chiurazzi, N. Crawford, F. Daniel-Vedele, L. David, R. Dickstein, E. Fernandez, B. Forde, W. Gassmann, D. Geiger, A. Gojon, J.M. Gong, B.A. Halkier, J.M. Harris, R. Hedrich, A.M. Limami, D. Rentsh, M. Seo, L.F. Tsay, M. Zhang, G. Coruzzi, and B. Lacombe, 2014: A unified nomemclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. Trends in Plant Sci 19, 5-9.

Li, X.P., X.Q. Zhao, X. He, G.Y. Zhao, B. Li, D.C. Liu, A.M. Zhang, X.Y. Zhang XY, Y.P. Tong, and Z.S. Li, 2010: Haplotype analysis of the genes encoding glutamine synthetase plastic isoforms and their association with nitrogen-use- and yield-related traits in bread wheat. New phytologist 189, 449-458.

Liu, Y., Z. He, R. Appels, and X. Xia, 2012: Functional markers in wheat: current status and future prospects. Theor. Appl. Genet. 125, 1-10.

Lizana, X.C., and D.F. Calderini, 2013: Yield and grain quality of wheat in response to increased temperatures at key periods for grain number and grain weight determination: considerations for the climatic change scenarios of Chile. J. Agric. Sci. 151, 209–221.

Long, S.P., X.G. Zhu, S.L. Naidu, and D.R. Ort, 2006: Can improvement in photosynthesis increase crop yields? Plant Cell Environ. 29, 315-330.

Longin, C.F.H., J. Mühleisen, H.P. Maurer, H. Zhang, M. Gowda, and J.C. Reif, 2012: Hybrid breeding in autogamous cereals. Theoretical and Applied Genetics 125:1087-1096.

López-Castañeda, C., R.A. Richards, G.D. Farquhar, and R.E. Williamson, 1996: Seed and seedling characteristics contributing to variation in early vigor among temperate cereals. Crop Sci 36, 1257-1266.

Loqué, D., and N. von Wirén, 2004: Regulatory levels for the transport of ammonium in plant roots. J Exp Bot 55, 1293-1305.

Ludewig, U., B. Neuhäuser, and M. Dynowski, 2007: Molecular mechanisms of ammonium transport and accumulation in plants. FEBS Lett 581, 2301-2308.

Lynch, J.P. 2007. Roots of the second green revolution. Austr. J. Bot. 55: 493-512.

Ma, Z., T.C. Walk, A. Marcus, and J.P. Lynch, 2001: Morphological synergism in root hair length, density, initiation and geometry for phosphorus acquisition in Arabidopsis thaliana: a modeling approach. Plant Soil 236, 221-235.

Maketon, C., A. M. Fortuna, and P. A. Okubara, 2012: Cultivar-dependent transcript accumulation in wheat roots colonized by Pseudomonas fluorescens Q8r1-96 wild type and mutant strains. Biol. Control 60, 216-224.

Malhi, S.S., and R.L. Lemke, 2013: Effectiveness of seedrow-placed N with polymer-coated and NBPT-treated urea for canola and wheat. J Plant Nutr 36, 2205-2224.

Manske, G. G. B., R. K. Behl, A. B. Luttger, and P. L. G. Vlek, 2000: Enhancement of mycorrhizal infection, nutrient efficiency and plant growth by Azotobacter in wheat: Evidence of varietal effects. In: N. Narula (ed). Azotobacter in Sustainable Agriculture, 136-147. CBS Publishers, New Delhi, India.

Mantelin, S., and B. Touraine, 2004: Plant growth promoting bacteria and nitrate availability: impacts on root development and nitrate uptake. J Exp Bot 394, 27-34.

Mantelin, S., G. Desbrosses, M. Larcher, T. J. Tranbarger, J. C. Cleyet-Marel, and B. Touraine, 2006: Nitrate-dependent control of root architecture and N nutrition are altered by a plant growth-promoting Phyllobacterium sp. Planta 223, 591-603.

Marshner, H., 1995: Mineral Nutrition of higher plants. London Academic Press, London.

Martre, P., J.R. Porter, P.D. Jamieson, and E. Triboï, 2003: Modeling grain nitrogen accumulation and protein composition to understand the Sink/Source regulations of nitrogen remobilization for wheat. Plant Physiol. 133, 1959–1967.

Martre, P., P.D. Jamieson, M.A. Semenov, R.F. Zyskowski, J.R. Porter, and E. Triboi, 2006: Modelling protein content and composition in relation to crop nitrogen dynamics for wheat. Europ. J. Agron. 25, 138-154.

Mathesius, U., 2009: Comparative proteomic studies of root–microbe interactions. J. Proteomics 72, 353-366.

Mazzola, M., D. L. Funnell, and J. M. Raaijmakers, 2004: Wheat cultivar-specific selection of 2,4-diacetylphloroglucinol-producing fluorescent Pseudomonas species from resident soil populations. Microb. Ecol. 48, 338-348.

McCullough, D.E. and L.A. Hunt, 1993: Mature tissue and crop canopy respiratory characteristics of rye, triticale and wheat. Ann. Bot. 72, 269-282.

Mcintyre, C.L., R.E. Casu, A. Rattey, M.F. Dreccer, J.W. Kam, A.F. van Herwaarden, R. Shorter, G.P. Xue, 2011: Linked gene networks involved in nitrogen and carbon metabolism and levels of water-soluble carbohydrate accumulation in wheat stems. Funct Integr Genomics 11:585-597.

Meynard, J.M., 1987: L'analyse de l'élaboration du rendement dans les essais de fertilisation azotée. Perspectives Agricoles 115, 76-83.

Miller, A.J., and S.J. Smith, 1996: Nitrate transport and compartimentation in cereal root cells. J Exp Bot 47, 843-854.

Miller, A.J., X. Fan, M. Orsel, S.J. Smith, and D.M. Wells, 2007: Nitrate transport and signaling. J Exp Bot 58, 2297-2306.

Moënne-Loccoz, Y., P. Mavingui, C. Combes, P. Normand, and C. Steinberg, 2014: Microorganisms and biotic interactions. In: J. C. Bertrand, P. Caumette, P. Lebaron, and P. Normand (eds). Environmental Microbiology: Fundamentals and Applications. Springer, Dordrecht, The Netherlands (In Press).

Molina-Favero, C., C. M. Creus, M. Simontacchi, S. Puntarulo, and L. Lamattina, 2008: Aerobic nitric oxide production by Azospirillum brasilense Sp245 and its influence on root architecture in tomato. Mol. Plant-Microbe Interact. 21, 1001-1009.

Moll, R.H., E.J. Kamprath, and W.A. Jackson, 1982: Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization. Agron 74, 562-564.

Monaghan, J.M., J.W. Snape, A.J.S. Chojecki, and P.S. Kettlewell, 2001: The use of grain

protein deviation for identifying wheat cultivars with high grain protein concentration and yield. Euphytica 122, 309e317.

Monneveux, P., P.H. Zaidi and C. Sanchez, 2005: Population density and low nitrogen affects yield-associated traits in tropical maize. Crop Science 45, 535–545.

Moreau, D., V. Allard, O. Gaju, J. Le Gouis, M.J. Foulkes, and P. Martre, 2012: Acclimation of leaf nitrogen to vertical light gradient at anthesis in wheat is a whole-plant process that scales with the size of the canopy Plant Physiol 160, 1479-1490.

Morgan, C.L., R.B. Austin, M.A. Ford, J. Bingham, W.J. Angus, and S. Chowdhury 1989: An evaluation of F1 hybrid winter wheat genotypes produced a chemical hybridizing agent. Journal of Agricultural Science, Cambridge 112:143-149.

Moubayidin, L., R. Di Mambro, and S. Sabatini, 2009: Cytokinin-auxin crosstalk. Trends Plant Sci. 14, 557-562.

Mühleisen, J., H.P. Piepho, H.P. Maurer, C.F.H. Longin, and J.C. Reif 2014: Yield stability of hybrids versus lines in wheat, barley and triticale. Theoretical and Applied Genetics 127:309-316.

Munier-Jolain, N., and C. Salon, 2005 : Are the carbon costs of seed production related to the quantitative and qualitative performance? An appraisal for legumes and other crops. Plant, Cell and Environment 28, 1388-1395.

Murchie, E.H., M. Pinto, and P. Horton, 2009: Agriculture and the new challenges for Photosynthesis Research. New Phytol. 181, 532-52

Muurinen, S., G.A. Slafer, and P. Peltonen Sainio, 2006: Breeding effects on Nitrogen Use Efficiency of Spring Cereals under northern Conditions. Crop Sci 46, 561-568.

Nacry, P., E. Bouguyon, and A. Gojon, 2013: Nitrogen acquisition by roots: physiological and developmental mechanisms ensuring plant adaptation to fluctuating resources. Plant Soil 370, 1-29.

Naruoka, Y., J.D. Sherman, S.P. Lanning, N.K. Blake, J.M. Martin, and L.E. Talbert, 2012: Genetic analysis of green leaf duration in spring wheat. Crop Sci. 52, 99-109.

Nasholm, T., K. Kielland, and U. Ganeteg, 2009: Uptake of organic nitrogen by plants. New Phytol 182, 31-48.

Neiverth, A., S. Delai, D. M. Garcia, K. Saatkamp, E. M. de Souza, F. de Oliveira Pedrosa, V. F. Guimarães, M. F. dos Santos, E. C. G. Vendruscolo, and A. C. T. da Costa, 2014: Performance of different wheat genotypes inoculated with the plant growth promoting bacterium Herbaspirillum seropedicae. Eur. J. Soil Biol. 64, 1-5.

Nelson, D. R., and P. M. Mele, 2006: The impact of crop residue amendments and lime on microbial community structure and nitrogen-fixing bacteria in the wheat rhizosphere. Soil Res. 44, 319-329.

Nguyen, C., 2003: Rhizodeposition of organic C by plants: mechanisms and controls. Agronomie 23, 375-396.

Nieder, R., D.K. Benbi, and H.W. Scherer, 2011: Fixation and defixation of ammonium in soils: a review. Biol Fertil Soils 47, 1-14.

O'Toole, J. C., and W.L. Bland, 1987: Genotypic variation in crop plant-root systems. Adv. Agron. 41, 91-145.

Ortiz, R., H.J. Braun, J. Crossa, J .H. Crouch, G. Davenport, J. Dixon, S. Dreisigacker, E. Duveiller, Z. He, J. Huerta, A.K. Joshi, M. Kishii, P. Kosina, Y. Manes, M. Mezzalama, A. Morgounov, J. Murakami, J. Nicol, G. Ortiz Ferrara, I. Ortiz-Monasterio, T.S. Payne, R.J. Pena, M.P. Reynolds, K.D. Sayre, R.C. Sharma, R.P. Singh, J. Wang, M. Warburton, H. Wu and M. Iwanaga, 2008: Wheat genetic resources enhancement by the International Maize and Wheat Improvement Center (CIMMYT). Genet Resour Crop Evol 55: 1095-1140.

Ortíz-Castro, R., H. A. Contreras-Cornejo, L. Macías-Rodríguez, and J. López-Bucio, 2009: The role of microbial signals in plant growth and development. Plant Signal Behav. 4, 701-712.

Ortiz-Monasterio, I., K.D. Sayre, S. Rajaram, and M. McMahon, 1997: Genetic progress in wheat yield and nitrogen use efficiency under four N rates. Crop Sci 37, 898-904.

Oury, F.X., P. Brabant, P. Pluchard, P. Bérard, and M. Rousset, 1994: Une étude de la qualité des blé hybrides à travers différents tests technologiques. Agronomie 14:377-385.

Oury, F.X., E. Triboï, P. Bérard, J.L. Ollier, and M. Rousset, 1995: Etude des flux de carbone et d'azote chez des blés hybrides et leurs parents, pendant la période de remplissage du grain. Agronomie 15:193-204.

Oury, F.-X., P. Bérard, M. Brancourt-Hulmel, C. Depatureaux, G. Doussinault, N. Galic, A. Giraud, E. Heumez, C. Lecomte, P. Pluchard, B. Rolland, M. Rousset, and M. Trottet, 2003: Yield and grain protein concentration in bread wheat: a review and a study of multi-annual data from a French breeding program. J. Genet. Breeding 57, 59-68.

Oury, F.X., and C. Godin, 2007: Yield and grain protein concentration in bread wheat: how to use the negative relationship between the two characters to identify favourable genotypes? Euphytica 157, 45–57.

Pace, G.M., and P.R. McClure, 1986: Comparison of nitrate uptake kinetic parameters across maize inbred lines. J Plant Nutr 8, 1095-1111.

Parry, M.A.J., P.J. Andralojc, R.A.C. Mitchell, P.J. Madgwick, and A.J. Keys, 2003: Manipulation of Rubisco: its amount, activity, function and regulation. J. Exp. Bot. 54, 1321-1333.

Parry, M.A.J., M. Reynolds, M.E. Salvucci, C. Raines, P.J. Andralojc, X.G. Zhu, G.D. Price, A.G. Condon, and R. Furbank, 2011: Raising Yield Potential of Wheat: (II) Increasing photosynthetic capacity and efficiency. J. Exp. Bot. 62, 453-468.

Pask, A.J.D., 2009: Optimising nitrogen storage in wheat canopies for genetic reduction in fertiliser nitrogen inputs. Ph.D. Thesis, University of Nottingham, UK.

Pask, A.J.D, R. Sylvester-Bradley, P.D. Jamieson, and M.J. Foulkes, 2012: Quantifying how wheat crops accumulate and use N during growth. Field Crops Res 126, 104-118.

Passioura, J., 1977: Grain yield, harvest index, and water use of wheat. Journal of the Australian Institute of Agricultural Science 43, 117-120.

Pathak, R.R., S. Lochab, and N. Raghuram, 2011: Plant Systems | Improving Plant Nitrogen-Use Efficiency. In: Murray Moo-Young (ed.), Comprehensive Biotechnology, Second Edition, volume 4, pp. 209-218. Elsevier.

Perezin, M., M. Corbellini, M. Accerbi, P. Vaccino, and B. Borghi 1998: Bread wheat : F1 hybrid performance and parental diversity estoimates using molecular markers. Euphytica 100, 273-279.

Perin, L., L. Martínez-Aguilar, R. Castro-González, P. Estrada-de los Santos, T. Cabellos-Avelar, H. V. Guedes, V. M. Reis, and J. Caballero-Mellado, 2006: Diazotrophic Burkholderia species associated with field-grown maize and sugarcane. Appl. Environ. Microbiol. 72, 3103-3110.

Phillips, D. A., T. C. Fox, M. D. King, T. V. Bhuvaneswari, and L. R. Teuber, 2004: Microbial products trigger amino acid exudation from plant roots. Plant Physiol. 136, 2887-2894.

Plett, D., J. Toubia, T. Garnett, M. Tester, B.N. Kaiser, and U. Baumann, 2010: Dichotomy in the NRT gene family of dicots and grass species. Plos One 5, e15289.

Pothier, J. F., C. Prigent-Combaret, J. Haurat, Y. Moënne-Loccoz, and F. Wisniewski-Dyé, 2008: Duplication of plasmid-borne nitrite reductase gene nirK in the wheat-associated plant growth-promoting rhizobacterium Azospirillum brasilense Sp245. Mol. Plant-Microbe Interact. 21, 831-842.

Prasad, P.V.V., P.Q. Craufurd, and R.J. Summerfield, 1999: Fruit number in relation to pollen production and viability in groundnut exposed to short episodes of heat stress. Ann. Bot. 84, 381-386.

Price, A.H, J. Townend, M.P. Jones, A. Audebert, and B. Courtois, 2002: Mapping QTLs associated with drought avoidance in upland rice grown in the Philippines and West Africa. Plant Mol. Biol. 48, 683-695.

Prigent-Combaret, C., D. Blaha, J. F. Pothier, L. Vial, M. A. Poirier, F. Wisniewski-Dyé, and Y. Moënne-Loccoz, 2008: Physical organization and phylogenetic analysis of acdR as leucine-responsive regulator of the 1-aminocyclopropane-1-carboxylate deaminase gene acdS in phytobeneficial Azospirillum lipoferum 4B and other Proteobacteria. FEMS Microbiol. Ecol. 65, 202-219.

Przystalski, M., A. Osman, E.M. Thiemt, B. Rolland, L. Ericsson, H. Østergård, L. Levy, M. Wolfe, A. Büschse, H.-P. Piepho, and P. Krajewski, 2008: Comparing the performance of cereals varieties in organic and non-organic cropping systems in different European countries. Euphytica 163, 417–433.

Ravel, C., P. Martre, I. Romeuf, M. Dardevet, R. El-Malki, J. Bordes, N. Duchateau, D. Brunel, F. Balfourier and G. Charmet, 2009: Nucleotide polymorphism in the wheat transcriptional activator Spa influences its pattern of expression and has pleiotropic effects on grain protein composition, dough viscoelasticity, and grain hardness. Plant Physiology 151: 2133-2144.

Rawluk, C.D.L., G.J. Racz, and C.A. Grant, 2000: Uptake of foliar or soil application of 15N-labelled urea solution at anthesis and its effect on wheat grain yield and protein. Can J Plant Sci 80, 331-334.

Reardon, C. L., H. T. Gollany, and S. B. Wuest, 2014: Diazotroph community structure and abundance in wheat–fallow and wheat–pea crop rotations. Soil Biol. Biochem. 69, 406-412.

Rebetzke, G.J. and R.A. Richards, 1999: Genetic improvement of early vigour in wheat Aust J Agric Res 50, 291-301.

Reynolds, M.P., M. Van Ginkel, and J.M. Ribaut, 2000: Avenues for genetic modification of radiation use efficiency in wheat. J Exp Bot 51, 459-473.

Reynolds, M.P., F. Dreccer, and R. Trethowan, 2007: Drought-adaptive traits derived from wheat wild relatives and landraces. J. Exp. Bot. 58, 177-186.

Reynolds, M.P., Y. Manes, A Izanloo, and P. Langridge, 2009: Phenotyping for physiological breeding and gene discovery in wheat. Ann. Appl. Biol. 155, 309-320.

Richardson, A. E., J. M. Baréa, A. M. McNeill, and C. Prigent-Combaret, 2009: Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. Plant Soil 321, 305-339.

Robertson, G.P., and P.M. Vitousek, 2009: Nitrogen in Agriculture: Balancing the Cost of an Essential Resource. Ann. Rev. Envir. Res. 34, 97-125.

Robinson, D. 2001: Root proliferation, nitrate inflow and their carbon costs during nitrogen capture by competing plants in patchy soil. Plant Soil 232, 41-50.

Ruuska, S.A., D.C. Lewis, G. Kennedy, R.T. Furbank, C.L.D. Jenkins, and L.M. Tabe, 2008: Large scale transcriptome analysis of the effects of nitrogen nutrition on accumulation of stem carbohydrate reserves in reproductive stage wheat. Plant Mol Biol 66, 15-32.

Sadras, V. O., and G. Lemaire, 2014: Quantifying crop nitrogen status for comparisons of agronomic practices and genotypes. Field Crops Research 164, 54-64.

Sadras, V. O., and R.A. Richards, 2014: Improvement of crop yield in dry environments: benchmarks, levels of organisation and the role of nitrogen. Journal of experimental botany 65, 1981-1995.

Sarcevic, H., K. Jukic, I. Ikic, and A. Lovric, 2014: Estimation of quantitative genetic parameters for grain yield and quality in winter wheat under high and low nitrogen fertilization. Euphytica 199, 57-67.

Sakakibara, Y., H. Kimura, A. Iwamura, T. Saitoh, T. Ikegami, G. Kurisu, and T. Hase, 2012: A new structural insight into differential interaction of cyanobacterial and plant ferredoxins with nitrite reductase as revealed by NMR and X-ray crystallographic studies. J Biochem 151, 483-492.

Semenov, M.A., P.D. Jamieson, and P. Martre, 2007: Deconvoluting Nitrogen Use Efficiency in Wheat: A Simulation Study. Europ J Agron 26, 283-294.

Sétif, P., M. Hirasawa, N. Cassan, B. Lagoutte, J.N. Tripathy, D.B. Knaff, 2009: New insights into the catalytic cycle of plant nitrite reductase. Electron transfer kinetics and charge storage. Biochemistry 48, 2828-2838.

Sharma, S., S. Xu, B. Ehdaie, A. Hoops, T.J. Close, A.J. Lukaszewski, and J.G. Waines, 2011: Dissection of QTL effects for root traits using a chromosome arm-specific mapping population in bread wheat. Theor. Appl. Genet. 122, 759–769.

Shrawat, A.K., R.T. Carroll, M. DePauw, G.J. Taylor, and A.G. Good, 2008: Genetic engineering of improved nitrogen use efficiency in rice by the

tissue-specific expression of alanine aminotransferase. Plant Biotech. J. 6, 337-345.

Shearman, V.J., R. Sylvester-Bradley, and M.J. Foulkes, 2005: Physiological processes associated with wheat yield progress in the UK. Crop Sci 45, 175-185.

Shewry, P.R., and N.G. Halford, 2002: Cereal seed storage proteins: structures, properties and role in grain utilization. Journal of Experimental Botany 53, 947–958.

Sidiqui, M.Y., A.D.M. Glass, T.J. Ruth, and T.W. Rufty, 1990: Studies of the uptake of nitrate in barley. Plant Physiol 93, 1426-1432.

Silva, S.A., F.I.F. de Carvalho, V.R. Caetano, A.C. Oliveira, J.L.M. Coimbra, N.J.S. Vasconcellos, and C. Lorencetti, 2000: Genetic basis of stay-green trait in bread wheat. J. New Seeds 2, 55-68.

Simmonds, N.W., 1995: The relation between yield and protein in cereal grain. J Sci Food Agric 67, 309.

Sinclair, T.R., and T. Horie, 1989: Leaf nitrogen, photosynthesis, and crop radiation use efficiency: a review. Crop Sci. 29, 90-98.

Sinclair, T.R., and J. Amir, 1992: A model to assess nitrogen limitations on the growth and yield of spring wheat. Field Crops Res. 30, 63-78.

SØgaard, R., M. Alstertfjord, N. MacAulay, and T. Zeuthen, 2009: Ammonium ion transport by the AMT/Rh homolog TaAMT1;1 is stimulated by acidic pH. Pflugers Arch 458, 733-743.

Srivastava, S., V. Chaudhry, A. Mishra, P. S. Chauhan, A. Rehman, A. Yadav, N. Tuteja, and C. S. Nautiyal, 2012: Gene expression profiling through microarray analysis in Arabidopsis thaliana colonized by Pseudomonas putida MTCC5279, a plant growth promoting rhizobacterium. Plant Signal. Behav. 7, 235-245.

Subbarao, G.V., T. Ban, M. Kishii, O. Ito, H. Samejima, H.Y. Wang, S.J. Pearse, S. Gopalakrishnan, K. Nakahara, A.K.M. Zakir Hossain, H.Tsujimoto, and W.L. Berry, 2007: Can biological nitrification inhibition (BNI) genes from perennial *Leymus racemosus* (Triticeae) combat nitrification in wheat farming? Plant and Soil 299: 55-64.

Sun, J.J., Y. Guo, G.Z. Zhang, M.G. Gao, G.H. Zhang, F.M. Kong, Y. Zhao, and S.S. Li, 2013: QTL mapping for seedling traits under different nitrogen forms in wheat. Euphytica 191, 317-331.

Suzuki, A., and D.B. Knaff, 2005: Glutamate synthase: Structural, mechanistic and regulatory properties, and role in the amino acid metabolism. Photosynth Res 83, 191-217.

Sylvester-Bradley, R., D.T. Stokes, R.K. Scott and V.B.A Willington, 1990: A physiological analysis of the diminishing responses of winter wheat to applied nitrogen. 2. Evidence. Aspects Appl Biol 25, 289-300.

Sylvester-Bradley, R., and D.R. Kindred, 2009: Analysing nitrogen responses of cereals to prioritize routes to the improvement of nitrogen use efficiency. Journal of experimental botany 60, 1939-1951.

Tardieu, F., 2013: Plant response to environmental conditions: assessing potential production, water demand, and negative effects of water deficit. Frontiers in Physiology 4, 17

Tenea, G.N., F. Cordeiro Raposo, and A. Maquet, 2012: Comparative Transcriptome Profiling in Winter Wheat Grown under Different Agricultural Practices. J Agric Food Chem 60, 10970-8.

Tercé-Laforgue, T., M. Bedu, C. Dargel-Graffin, F. Dubois, Y. Gibon, F.M. Restivo, and B. Hirel, 2013: Resolving the role of plant glutamate dehydrogenase: II. Physiological Characterization of plants overexpressing individually or simultaneously the two enzyme subunits. Plant Cell Physiol 54,1634-1647.

Tester, M., and P. Langridge, 2010: Breeding technologies to increase crop production in a changing world. Science 327, 818-822.

Tétard-Jones, C., P.N. Shotton, L. Rempelos, J. Cooper, M. Eyre, C.H. Orr, C. Leifert, and A.M.R. Gatehouse, 2013: Quantitative proteomics to study the response of wheat to contrasting fertilisation regimes. Mol Breeding 31:379-393.

Thomas, H. and C.M. Smart, 1993: Crops that stay green. Annals of Applied Biology, 123, 193–219.

Thomsen, H.C., D. Erikson, I.S. Møller, and J.K. Schoerring, 2014: Cytosolic glutamine synthetase: a target for improvement of crop nitrogen use efficiency. Trends Pl Sci 19, 656-663.

Thorne, G.N., W. Pearman, W. Day, and A.D. Todd, 1988: Estimation of radiation interception by winter wheat from measurements of leaf area. J Agric Sci Camb 110, 101-108.

Triboï, E., and J. Ollier, 1991: Kinetic and role of carbon and nitrogen stored in stem on 21 wheat genotypes. Agronomie 11, 239-246.

Uauy, C., A. Distelfeld, T. Fahima, A. Blechl, and J. Dubcovsky, 2006: A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314, 1298-1301.

Upadhyay, S. K., D. P. Singh, and R. Saikia, 2009: Genetic diversity of plant growth promoting rhizobacteria isolated from rhizospheric soil of wheat under saline condition. Curr. Microbiol. 59, 489-496.

Van Oosterom, E.J., A.K. Borrell, S.C. Chapman, I.J. Broad, and G.L. Hammer, 2010a: Functional dynamics of the nitrogen balance of sorghum: I. N demand of vegetative plant parts. Field Crop Res. 115, 19-28.

Van Oosterom, E.J., S.C. Chapman, A.K. Borrell, I.J. Broad, and G.L. Hammer, 2010b: Functional dynamics of the nitrogen balance of sorghum: II. Grain filling period. Field Crop Res. 115, 29-38.

Venieraki, A., M. Dimou, P. Pergalis, I. Kefalogianni, I. Chatzipavlidis, and P. Katinakis, 2011: The genetic diversity of culturable nitrogen-fixing bacteria in the rhizosphere of wheat. Microb. Ecol. 61, 277-285.

Veresoglou, S. D., and G. Menexes, 2010: Impact of inoculation with Azospirillum spp. on growth properties and seed yield of wheat: a meta-analysis of studies in the ISI Web of Science from 1981 to 2008. Plant Soil 337, 469-480.

Verma, V., M.J. Foulkes, P. Caligari, R. Sylvester-Bradley, and J. Snape, 2004: Mapping quantitative trait loci for flag leaf senescence as a yield determinant in winter wheat under optimal and drought-stressed environments. Euphytica 135, 255-263.

Von Wirén, N., S. Gazzarini, and W.B. Frommer, 1997: Regulation of mineral uptake in plants. Plant and Soil 196, 191-199.

Von Wirén, N., and M. Merrick, 2004: Regulation and function of ammonium carriers in bacteria, fungi and plants. Top Curr Genet 9, 95-120.

Walk, T., R. Jaramillo, and J.P. Lynch, 2006: Architectural tradeoffs between adventitious and basal roots for phosphorus acquisition. Plant and Soil 279, 347-366.

Walker, V., O. Couillerot, A. Von Felten, F. Bellvert, J. Jansa, M. Maurhofer, R. Bally, Y. Moënne-Loccoz, and G. Comte, 2012: Variation of secondary metabolite levels in maize seedling roots induced by inoculation with Azospirillum, Pseudomonas and Glomus consortium under field conditions. Plant Soil 356,151-163.

Wang, M.Y., M.Y. Siddiqi, T.J. Ruth, and A.D.M. Glass, 1993: Ammonium uptake by rice roots (II. Kinetics of $13NH_4^+$ influx across the plasmalemma). Plant Physiol 103, 1259-1267.

Wan, Y., P.R. Shewry, and M.J. Hawkesford , 2013: A novel family of γ-gliadin genes are highly regulated by nitrogen supply in developing wheat grain. J Exp Bot 64, 161-168.

Wang,Y.Y., P.K. Hsu, and Y.F. Tsay, 2012: Uptake, allocation and signaling of nitrate. Trends in Plant Sci 17, 458-467.

Wang, Z.K., Z.F. Ni, H.L. Wu, X.L. Nie, and Q.X. Sun, 2006: Heterosis in root development and differential gene expression between hybrids and their parental inbreds in wheat (Triticum aestivum L.). Theoretical and Applied Genetics 113:1283-1294.

Wang, W.H., B. Köhler, F.Q. Cao, G.W. Liu, Y.Y. Gong YY, S. Sheng, Q.C. Song, X.Y. Cheng, T. Garnett, M. Okamoto, R. Qin, B. Mueller-Rober, M. Tester; and L.H. Liu, 2012: Rice DUR3 mediates high-affinity urea transport and plays an effective role in improvement of urea acquisition and utilization when expressed in *Arabidopsis*. New Phytol 193, 432-444.

Witte, C.P., 2011: Urea metabolism in plants. Plant Sci 180, 431-438.

Wu, H., T. Haig, J. Pratley, D. Lemerle, and M. An, 2001: Allelochemicals in wheat (Triticum aestivum L.): cultivar difference in the exudation of phenolic acids. J. Agr. Food Chem 49, 3742-3745.

Xu, G., X. Fan, and T. Miller, 2012: Plant nitrogen assimilation and use efficiency. Ann Rev Plant Biol 63, 153-182.

Xu, Y., R. Wang, Y. Tong, H. Zhao, Q. Xie, D. Liu, A. Zhang, B. Li, H. Xu, and D. An, 2013: Mapping QTL for yield and nitrogen-related traits in wheat: influence of nitrogen and phosphorus fertilization on QTL expression. Theor Appl Genet 127, 59-72.

Yamaya, T., and M. Kusano, 2014: Evidence supporting distinct functions of three cytosolic glutamine synthetases and two NADH-glutamate synthases in rice. J Exp Bot 65, 5519-5525.

Yin, L.P., P. Li, B. Wen, D. Taylor, and J.O. Berry, 2007: Characterization of a high-affinity nitrate system transporter gene (*TaNRT2*.1) from wheat roots and its evolutionary relationship to other *NRT2* genes. Plant Sci 172, 621-631.

Zheng, B.S., J. Le Gouis, M. Leflon, W.Y. Rong, A. Laperche, and M. Brancourt-Hulmel, 2010: Using probe genotypes to dissect QTL × environment interactions for grain yield components in winter wheat. Theor Appl Genet 121, 1501-1517.

Zhu, X.G., S.P. Long, and D.R. Ort, 2010: Improving photosynthetic efficiency for greater yield. Annu Rev Plant Biol. 61, 235-261.

EXTENT AND LIMITATION OF THE DATASET

During this PhD thesis, we used a dataset composed of eight experiments defined as a combination of locations, years, and nitrogen regimes. And a total of 225 varieties were evaluated in a well balanced design. In each experiment, 18 environmental covariates were computed and 28 NUE-related traits were measured or calculated. A more exhaustive description is provided in the following parts of the manuscript and in annexes. Here, we will mainly describe the environmental variability (combination of year and location) and discuss about its consequences.

Field trials

All experiments were conducted in the North of France, which is the main wheat producing region of the country (Fig. 3A). Thus, we have to keep in mind that varieties coming from breeding stations located in this area may be favoured. In fact, this can create a confounding effect of genes determining regional adaptation which may be assessed as having additive effects. However, tested in a wider range of environments, they would have been assessed as interacting with the environments and/or not having any additive effects.

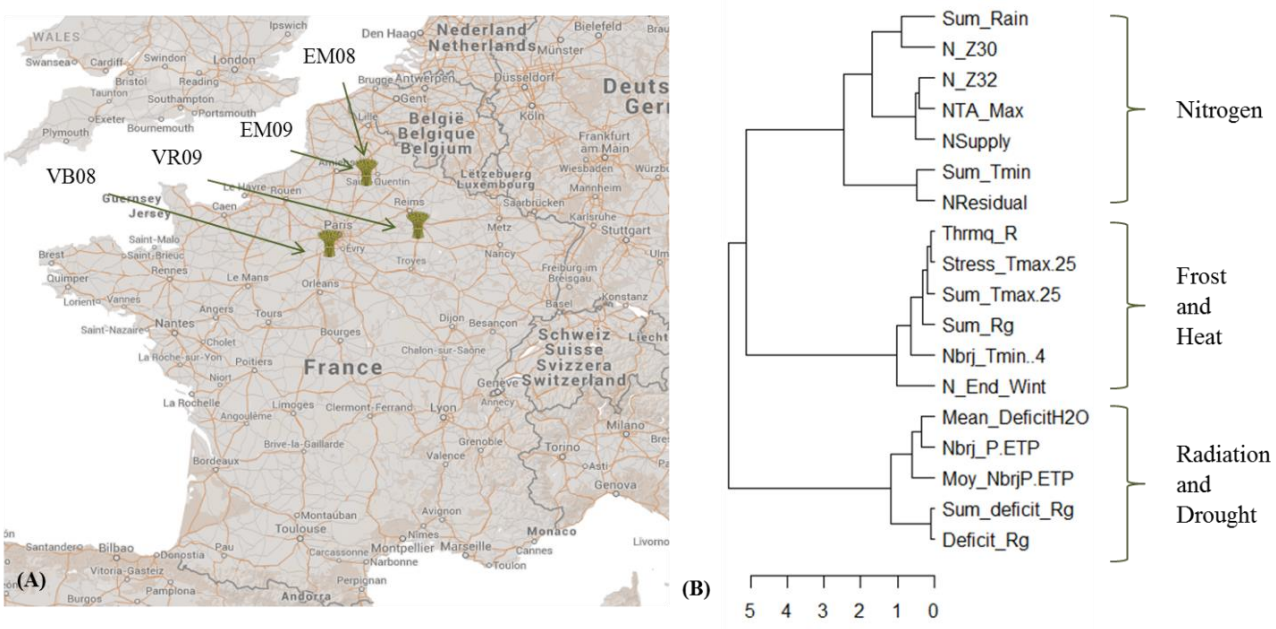


Figure 3: (A) Trial locations and (B) dendrogram of environmental covariates (from PCA analyses). The table used to perform the PCA is provided in Annexes (Annexes of Part IV). Clustering using PCA coordinates (hclust, method = ward).

Experiments were conducted during the 2007/2008 and 2008/2009 growing seasons. Looking at the specificity of these two growing seasons (Météo France information), we can estimate the range of

environmental stresses and thus, discuss the portability of our results. The winter of 2007/2008 had a high mean temperature and was the 10th warmest since the beginning of the XXth century. In contrast, the winter of 2008/2009 was the third coldest winter between 1990 and 2009. Spring 2008 was characterized by strong moisture, weak radiation, and warm temperatures particularly in May. Spring 2009 was dry, very sunny and even warmer making it the seventh warmest spring since 1900. Summer 2008 was unexceptional, while summer 2009 was dry and hot.

To conclude, these two growing seasons were really contrasted and embodied the main climatic variation occurring in the tested area. Thus, our dataset allowed for the study of a good variability of frost, spring drought and radiation stresses added to the on purpose applied N stresses. Our environmental covariates took into account these variations. Indeed, using principal component analysis, experiments first clustered by year and then by location with enhanced hydric and heat stresses for EM09 and VR09, respectively.

Regarding the diversity of occurring stresses, the main limitation of our dataset is its size. In fact, due to the reduced number of environments, stresses are not independent and effects can be confounded. For example, radiation and drought stresses are linked together as frost and heat stresses are (Fig. 3B). Applied N stresses are also linked to other environmental covariates. For exmaple, soil residual N is linked to the winter hardness (Sum_Tmin) and quantity of N applied at Z30 (1cm spike; N_Z30) is linked to the sum of rain (Sum_rain). Thus, frost stress may have enhanced mineralisation. Differences in N applied between experiments may be enhanced by precipitation which influences the availability of N to plant. But, it also means that varieties responses to N stresses could be in part confounded with varieties responses to frost and drought stresses.

Tested genotypes

Following the initiative of Arvalis institut du végétal, Biogemma decided in 2007 to focus on the elite registered variability and initiated physiological, agronomical and molecular characterization of this material through the building of a panel. Each year this panel was enriched by 20-30 new varieties. Meanwhile, the oldest varieties or the worst ones (commercially speaking) were removed. In this thesis, we worked on the 2007/2008 and 2008/2009 versions of this panel.

Our panel is composed of European elites released from 1969 to 2010 and selected in different European breeding programs. Thus, we are studying certain among of physiological and genetic diversity. Nevertheless, some of these varieties were selected to perform well where we tested them and others were selected for other environments. The main criterion for adaptation in wheat is earliness that can be approximated by flowering date. In our panel, the standard deviation in flowering date was seven days, which is significant. Consequently, we have to be aware of the previously mentioned confounding effect on adaptation genes and may use flowering date as a covariate in some analyses to take into account regional adaptation.

In our panel, physiological diversity can arise from selection effect as we used a historical panel. Consequently, we will have to check if the associated chromosomal regions are not fixed in the more recent

varieties. Otherwise, these results may not be useful in selection, which mostly uses elite × elite crosses. For traits that were not under selection pressure, it remains to be seen whether the diversity is sufficient in elite germplasm to actually start to select for them.

Regarding genetic diversity, using elite varieties instead of exotic ones may reduce the frequency of unusable loci due to a low minor allele frequency. This also means that numerous loci will not be polymorphic at all and their effects will not be assessed whether they are positive or negative. We can illustrate that with the use of the 90K gene-associated SNP chip developed using transcriptome sequencing of a broader genetic diversity (Wang et al. 2014). On the total number of SNP that were properly scored (36K on 90K), around 28% were monomorphic. Added to the 90K chip, we used SNP developed by Biogemma (30% of the total genotyping dataset). This dataset is not publically available to give a competitive advantage to Biogemma and contains SNP mainly located in candidate regions or genes. Thus, we may be more focused on particular regions. More generally, we mostly focused on genic variability. To conclude, we screened a genetic diversity which is reduced by our panel and partially biased by our selection of SNP.

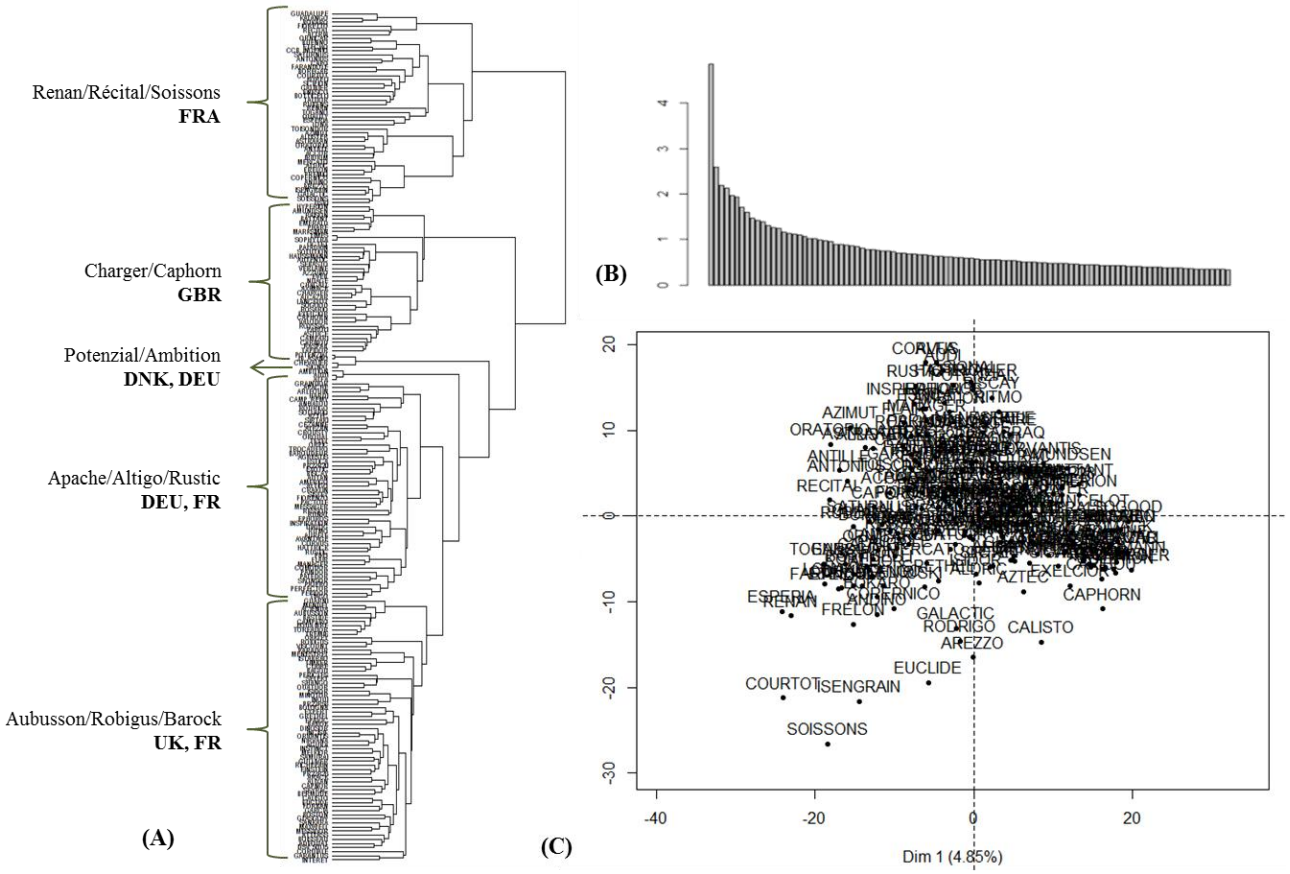


Figure 4: (A) Dendrogram of individuals and (B) percentage of variance explained by axis in the (C) principal component analysis. In the dendrogram, varieties are clustered using the kinship matrix (method Ward). PCA analysis was performed on the genotyping matrix (genotypes × SNP).

First, we should be aware of physiological and genetic variances as they are impacting statistical power in the linkage disequilibrium mapping methods that we used. Panel structure also impacts statistical power. In fact, phenotypic variance is only useful if it is not linked to the panel structure. In Europe, commercial lines can be re-used in concurrent breeding programs. Thus, European elite lines are not well structured even if varieties have a tendency to cluster by breeding companies and geographical origin (Fig. 4). In agreement to this, following Patterson et al. (2006) who developed a statistical method to test the significance of structure, we concluded that we did not have any significant structure in our panel. This absence of a strong structure is good news and may compensate a reduced phenotypic variance. However, structure studies are performed at the panel scale. At a smaller scale, varieties kinship is not uniform and should be taken into account. Moreover, wheat market is segmented in different classes of quality under the genetic determinism of a reduced number of genes having a huge influence on agronomic performances. And, this information may be "diluted" in the kinship matrix. Thus, quality classes may have to be used as a covariate in some analyses as flowering date.

Our dataset is obviously limited but allows for the study of varieties' responses to a wide range of environmental stresses. Moreover, using elite varieties, our results will completely be (i) in the scope of breeders working on winter wheat adapted to North West Europe and (ii) in the scope of Arvalis institut du vegetal, which mission is to advice farmers on cultural practice maximizing yield potential for a given variety. As previously mentioned, before looking at the genetic determinism of NUE related traits, the first question to answer is whether phenotypic diversity is sufficient in our panel. Moreover, past breeding effort can be analysed and discussed in order to better design the future one.

References

Météo France: http://www.meteofrance.fr/climat-passe-et-futur/bilans-climatiques/bilan-2008/bilan-de-lhiver-2007-2008 ; http://www.meteofrance.fr/bilan-de-lhiver-20082009

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. Plos Genet 2:e190 doi:10.1371/journal.pgen.0020190.

Wang S, Debbie W, Forrest K, Allen A, et al. (2014) Characterization of polyploidy wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotech J, 12:787-796.

PART II: ONCE UPON A TIME

Theoretical and Applied Genetics
December 2013, Volume 126, Issue 12, pp 3035-3048
Accepted: 6 September 2013

A MULTI-ENVIRONMENTAL STUDY OF RECENT BREEDING PROGRESS ON NITROGEN USE EFFICIENCY IN WHEAT (*T. AESTIVUM* L.)

Fabien Cormier[1] • Sébastien Faure[1] • Pierre Dubreuil[1] • Emmanuel Heumez[2] • Katia Beauchêne[3] • Stéphane Lafarge[1] • Sébastien Praud[1]✉ • Jacques Le Gouis[4]

[1]Biogemma, Centre de recherche de Chappes, Route d'Ennezat CS90126, 63720 Chappes, France. [2]INRA UE GCIE, 2 Chaussée Brunehaut Estrées-Mons BP50136, 80203 Péronne, France.[3]ARVALIS-Institut du végétal, IBP, Bât. 630, Université Paris Sud, Rue de Noetzlin, 91405 Orsay Cedex, France. [4]INRA, UBP UMR 1095, Génétique, Diversité et Ecophysiologie des Céréales, 24 Avenue des Landais, 63177 Aubière Cedex, France.
✉ sebastien.praud@biogemma.com

ABSTRACT: In a context where European agriculture practices have to deal with environmental concerns and nitrogen (N) fertiliser cost, nitrogen use efficiency (NUE) has to be improved. This study assessed genetic progress in winter wheat (*Triticum aestivum* L.) NUE. Two hundred and twenty-five European elite varieties were tested in four environments under two levels of N. Global genetic progress was assessed on additive genetic values and on genotype × N interaction, covering 25 years of European breeding. To avoid sampling bias, quality, precocity and plant height were added as covariates in the analyses when needed. Genotype × environment interactions were highly significant for all the traits studied to such an extent that no additive genetic effect was detected on N uptake. Genotype × N interactions were significant for yield, grain protein content (GPC), N concentration in straw, N utilisation, and NUE. Grain yield improvement (+0.45 % year^{-1}) was independent of the N treatment. GPC was stable, thus grain nitrogen yield was improved (+0.39 % year^{-1}). Genetic progress on N harvest index (+0.12 % year^{-1}) and on N concentration in straw (-0.52 % year^{-1}) possibly revealed improvement in N remobilisation. There has been an improvement of NUE additive genetic value (+0.33 % year^{-1}) linked to better N utilisation (+0.20 % year^{-1}). Improved yield stability was detected as a significant improvement of NUE in low compared to high N conditions. The application of these results to breeding programs is discussed.

ABBREVIATIONS

ADM_S, straw dry matter at maturity; BLUE, best linear unbiased estimator; BLUP, best linear unbiased predictor; E, environment; FLO, flowering date; G, genotype; GNY, grain nitrogen yield; GPC, grain protein content; GPD, grain protein deviation; GY, grain dry matter yield; HI, harvest index; HN, high nitrogen input; KS, kernel per spike; LN, low nitrogen input; LRT, likelihood ratio test; LSD; Fisher's least significant difference test; N, nitrogen; %N_S, straw nitrogen content at maturity; NHI, nitrogen harvest index; NSA, straw nitrogen per area; NTA, total nitrogen in plant at maturity; NUE, nitrogen use efficiency; NUE_Prot, nitrogen use to protein efficiency; NupE, nitrogen uptake; NutE, nitrogen utilisation efficiency; NutE_Prot; nitrogen utilisation to protein efficiency; P; P-value; PH, plant height; SA, spike per area; TKW, thousand kernel weight; YR, year of release

INTRODUCTION

Nitrogen (N) fertiliser accounted for the majority (77.4 %) of nutrients consumed in Europe on all crops in 2011 (ec.europa.eu/eurostat). Its increasing application has largely contributed to bread wheat (*Triticum aestivum L.*) yield rise during the second half of the twentieth century (Erisman et al. 2008). But the cost of N fertiliser production and application is increasing (Rothstein, 2007) and environmental concerns (Goulding, 2004) make it necessary to enhance crop nitrogen use efficiency (NUE).

Two strategies may be devised for NUE improvement: maintaining high yield when reducing N supply, and/or increasing yield at a constant N supply. The cost of N production, environmental pollution due to nitrate leaching (Pathak et al. 2011), and volatilisation of greenhouse gases require that wheat NUE should be improved at a lower N supply. But the situation is more complex since increasing world demand for grain (Bruinsma, 2009) means that increased production per unit area is the priority. Thus, the minimum N rate to maximise yield should be considered. End-use is also an important factor as breadmaking, feed, or biofuel wheat varieties have different protein content requirements (Bushuk, 1998; Shewry and Halford 2002). Moreover, for a given cultivar, the maximal grain protein concentration and the maximal yield are generally not obtained with the same fertilisation strategy, *i.e.* amount and application dates (Lopez-Bellido et al. 2006). We should also notice that both lodging (Ortiz-Monasterio et al. 1997a) and foliar disease (Olesen et al. 2003) risks increase with N fertilisation.

Moll et al. (1982) defined NUE as grain dry matter (GY) divided by available N from the soil and fertiliser. Improving NUE is a relevant challenge for winter wheat for which N recovery and NUE are estimated to be respectively around 65 % and 25 kg DM kg^{-1} N at high N input in Northern Europe (Sylvester-Bradley and Kindred 2009; Gaju et al. 2011). As an integrative trait, NUE is usually decomposed into two components: the uptake and utilisation efficiencies. Uptake efficiency characterizes the capacity to capture N from the soil: it is often computed as total nitrogen in the plant at harvest (NTA) divided by available N in the soil. Utilisation efficiency characterises the capacity to convert total plant nitrogen to grain dry matter (GY / NTA).

The identification of traits to improve NUE in wheat and the characterisation of their variability provide useful directions to breeders (e.g. Barraclough et al. 2010; Foulkes et al. 2009; Gaju et al. 2011). The first decision that breeders have to take is to choose the N level for which they want to breed. Indeed, in numerous studies which analysed

agronomic traits, significant genotype × N (G × N) interactions were detected (e.g. Le Gouis et al. 2000; Laperche et al. 2006a; Barraclough et al. 2010), meaning that variety behaviour differentially depends on N treatment. Quantifying G × N interactions is therefore crucial for efficient selection. Recent selection in Europe has been conducted mostly at high or optimum N levels so genetic progress achieved at lower N levels results from indirect selection. As G × N interactions have been shown to increase with N stress (Bänziger et al. 1997; Laperche et al. 2006a) the efficiency of indirect selection for a low N input (LN) environment resulting from direct selection in a higher N input (HN) environment can be highly variable (Atlin and Frey 1989; Ceccarelli et al. 1992; Sinebo et al. 2002; Brancourt-Hulmel et al. 2005).

Characterizing and quantifying recent genetic progress can also bring meaningful information to breeders. Many studies have been conducted on wheat yield genetic progress (e.g. for recent studies Brisson et al. 2010; Fischer et al. 2010; Oury et al. 2012; Graybosch and Peterson, 2012; Lopez et al. 2012; Green et al. 2012). The main conclusion from studies conducted at different N levels is that genetic progress occurred in both HN and LN conditions, but was higher at HN (Ortiz-Monasterio et al. 1997a; Brancourt-Hulmel et al. 2003; Guarda et al. 2004). Fewer studies have been published on the genetic progress for NUE and its components (Ortiz-Monasterio et al. 1997a; Guarda et al. 2004; Muurinen et al. 2006). Moreover, it is well known that a negative correlation between yield and protein content exists in wheat (Kibite and Evans 1984; Simmonds, 1995, Oury et al. 2003; Oury and Godin, 2007, Bogard et al. 2010). A yield increase may therefore lead to a decrease in protein content which could cause lower end-use quality (Ortiz-Monasterio et al., 1997b; Shewry, 2004). Thus the question of the genetic improvement in yield or NUE cannot be assessed independently of quality.

Two major approaches are used to assess genetic progress: (i) historical trial analyses and (ii) direct comparisons of old and modern varieties in the same environment. But these two approaches suffer from some limitations. (i) When historical trials are analysed, as genotypes are tested in different year × environment combinations, there is a need to take into account agroclimatic variation. This may induce bias as elimination of "year" effects is often based on variation from year to year of common controls leading to inadequate consideration of genotype × "year" interactions (e.g. Brisson et al. 2010; Oury et al. 2012; Graybosch and Peterson 2012). (ii) Direct comparisons of old and modern varieties are often limited by the experiment size (*e.g.* Brancourt-Hulmel et al. 2003; Guarda et al. 2004; Muurinen et al. 2006; Green et al. 2012) with few genotypes studied in few environments. This can cause sampling errors. Lopez et al. (2012) proposed to base genetic progress assessment only on the highest yielding variety per date of release but still with a quite low number of cultivars. Moreover, the period under study is usually spread out and includes major changes in plant height due to introduction of dwarfing alleles. Indeed, height decrease is one of the major sign of winter wheat genetic improvement between 1946 and 1992 in France (Brancourt-Hulmel et al. 2003) as well as other countries (eg Ortiz-Monasterio et al. 1997a; Austin, 1999). It is directly linked to NUE through an increase of lodging resistance and nitrogen partitioning (Hedden, 2003). Plant height is now

stabilised, therefore the question of recent genetic gain can be asked independently of this major physiological change using a large panel of recent cultivars grown in the same environments.

Our work aims to assess recent genetic progress in NUE and NUE-related traits in HN and LN environments. For this purpose, (i) we assessed the additive genetic and interactive variances for NUE and its components, and (ii) we estimated genetic progress made during the last 25 years for both additive genetic effects and for G × N interactions. For this, we analysed a multi-environment dataset of eight independent trials (four HN input and four LN input) where 225 registered winter wheat varieties were directly compared.

Table 1: Description of the experimental design where wheat genotypes were evaluated at high N level (HN) and low N level (LN). NTA_{max} corresponds to the 95th percentile of total nitrogen per area at maturity for all the genotypes present in the trial and is an estimate of N available (soil + fertiliser N).

Site x Season	Season	Location	Soil type	Genotypes tested	Residual soil N (kg N ha^{-1})	N supply [a] (kg N ha^{-1})		NTA_{max} (kg N ha^{-1})	
						HN	LN	HN	LN
EM08	07/08	Estrées-Mons (49.8N,3.03E)	Clay loam	206[b]	67	50+70+50	0+70+0	206	144
EM09	08/09			208[b]	30	50+50+50	0+50+0	241	111
VB08	07/08	Villiers le Bacle (48.7N,2.1E)	Clay loam	197	106	0+66.5+60	0+44+0	242	157
VR09	08/09	Vraux (49.0N,4.2E)	White Chalk	196	30	60+100+60	60+60+0	236	173

[a] *Nsupply: fertiliser supply at end of winter + at Z30 + at Z32.*

[b] *controls: Apache, Orvantis, Caphorn, and Soissons (2007/08) or Premio (2008/09)*

MATERIALS AND METHODS

Plant materials and field experiments

Two hundred and twenty five European elite varieties released from 1969 to 2010 were evaluated in four environments (Table 1) as a combination of two sites and two seasons (Suppl. data 1 and Suppl. data 2). VB08 and VR09 were conducted by Arvalis experimental units in Villier-le-Bâcle and Vraux. EM08 and EM09 were conducted by the INRA experimental unit in Estrées-Mons. Genotypes were ranked by heading date to limit competition, effects and distributed in eight blocks. At EM08 and EM09, an augmented design was used where four controls were repeated in each of the eight blocks. At VB08 and VR09 all varieties were repeated twice in a complete block design. Two nitrogen supply modalities were tested in each environment (Table 1). The high N (HN) treatment corresponds to common agricultural practice in the tested environments. The low N (LN) treatment corresponds to HN reduced by around 100 kg N ha^{-1}. Other crop inputs including weed, disease and pest control, and potassium, phosphate and sulphur fertilisers, were applied at sufficient levels to prevent them from limiting yield. Plant growth regulator was applied to limit lodging on all trials. A trial is defined as a combination of environment × N treatment (e.g. EM08_LN).

Phenotypic data

Plant height (PH) and the number of spikes per unit area (SA) were assessed on each plot except for VB08_LN where measurements were taken on only one replicate. Flowering date (FLO), thousand kernel weight (TKW), straw dry matter at maturity (ADM_S), straw nitrogen content at maturity (%N_S), grain dry matter (GY), and grain protein concentration (GPC) were measured on each plot in all trials. The number of kernel per spike (KS) was calculated as GY / (TKW × SA). Total nitrogen per unit area at maturity (NTA) was calculated as grain nitrogen yield [GNY = (GPC / 5.7) × GY] added to straw nitrogen per unit area (NSA = ADM_S × %N_S).

NUE was not calculated as proposed by Moll et al. (1982). Rather, considering that mineralisation, leaching and rain all impact on the estimation of available soil N (Hirel et al. 2007; Gaju et al. 2011; Bingham et al. 2012), in each trial total N available to plants was estimated as the 95th percentile of the NTA (NTA_{max}) (Table 1). Nitrogen use efficiency (NUE) was then estimated as GY divided by NTA_{max}. N uptake efficiency at maturity (NupE) was calculated as NTA divided by NTA_{max}. N utilisation efficiency (NutE) was calculated as GY divided by NTA. To illustrate the capacity of varieties to convert N into protein, N use efficiency for protein production (NUE_Prot = GPC / NTA_{max}) and N utilisation efficiency for protein production (NutE_Prot = GPC / NTA) were also computed. Harvest index (HI) was defined as the grain dry matter divided by the total dry matter [GY / (GY + ADM_S)]. N harvest index (NHI) at maturity was the amount of N in the grain compared to the total nitrogen in the plant (GNY / NTA). Grain protein deviation (GPD) was the deviation from the linear regression of GPC by GY in each trial (Monaghan et al. 2001).

In all trials, adjusted means were calculated using a linear model with varieties and blocks as fixed factors. This resulted in eight different datasets with 182 varieties in common. The other varieties were at least present in four trials. Adjusted means were then used in all the following analyses.

Mixed-model and variance decomposition

To P_{ijk}, the phenotype of genotype i (i=1… 225) in environment j (VB08, VR09, EM08, and EM09) with N treatment k (HN and LN), the following mixed-model was used:

$$P_{ijk} = \mu + N_k + E_j + E_j \times N_k + G_i + G_i \times E_j + G_i \times N_k + \varepsilon_{ijk} \quad (1)$$

and in the single N treatment analyses, the following reduced mixed-model was used:

$$P_{ij} = \mu + E_j + G_i + \varepsilon_{ij} \quad (2)$$

In both equations (1) and (2) μ is the general mean, N_k the fixed effect of N, E_j the random effect of the environment, $E_j \times N_k$ the environment × N level interaction, G_i the random additive effect of the variety. $G_i \times E_j$ and $G_i \times N_k$ are respectively effects for the variety x environment (G × E) interaction, and variety × N modality interaction (G × N). $\varepsilon_{ijk} \sim N(0, \sigma^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$ are residual error terms.

Fixed effects were tested using Wald tests. Variance components of random factors were tested

one by one using the likelihood ratio test (LRT) (Kendall and Stuart 1979), based on log-likelihood (Lmax) differences between the complete (1) and reduced models (1) without the tested factor.

LRT = - 2 × [log(Lmax full model) - log(Lmax reduced model)].

LRT is expected to be distributed as a χ^2 with degrees of freedom (*df*) as:

df = n_{PAR} full model – n_{PAR} reduced model
where n_{PAR} is the number of parameters.

The null hypothesis (no significant effect of the tested component) was rejected when LRT > χ^2 (*df*). In our case, *df* was 1 as it was assumed no genetic covariance among varieties nor covariance among the trials.

Heritability

Generalised heritability (h^2g) was calculated using the following formula developed by Cullis et al. (2006).

$h^2g = 1 - PEV / (2 \times \sigma^2g)$
where σ^2g is the genetic variance and PEV is the average pairwise prediction error variance of the genetic effects best linear predictions (BLUPs).

Effect of the year of registration

To test for genetic progress, G_i and $G_i \times N_k$ were calculated from equation (1) modified with G_i and $G_i \times N_k$ as fixed effects to avoid shrinkage issues. Effect of the year of release (YR) was assessed on additive genetic effect (G_i) and on the genotype × N level interaction term ($G_i \times N_k$) by variance analyses (ANOVA) in a linear model. These tests were also conducted with the quality classes, precocity, and plant height as covariates (Suppl. data 2). A complete model including all covariates was first computed but only significant covariates were kept in the final analyses. Quality and plant phenology (height and precocity) are correlated to the studied traits so using them as covariates to estimate genetic progress corrects for two potential errors. The first is an artificial evolution of the studied trait due to the non-homogeneous allocation of quality, precocity, or height among years, assuming that they would not have evolved during the period under study. Secondly, it also compensates the possible non-adaptation of varieties to the tested environments as in our panel varieties were selected for different European target environments.

The five quality classes used correspond to those of the National Association of French Millers: very high quality, high quality, good quality, biscuit quality, and other use. YR were found in the French (http://cat.geves.info/Page/ListeNationale) and the European catalogue of crop species (http://ec.europa.eu/food/plant/propagation/catalogues). Anthesis date and plant height best linear unbiased estimators (BLUEs) from the reviewed equation (1) were used as precocity and height covariates.

Only three varieties were released between 1969 and 1985. To avoid sampling bias these varieties were not included in the genetic progress analyses. In total, 195 European elite varieties for which quality and YR information were available were used to assess the genetic progress.

Software

Statistical analyses were performed using R.2.13.2 (The R development core-team 2012) and the ASReml-R package v3.0.1 (Butler et al. 2009; http://www.vsni.co.uk).

RESULTS

Grain yield and N efficiencies

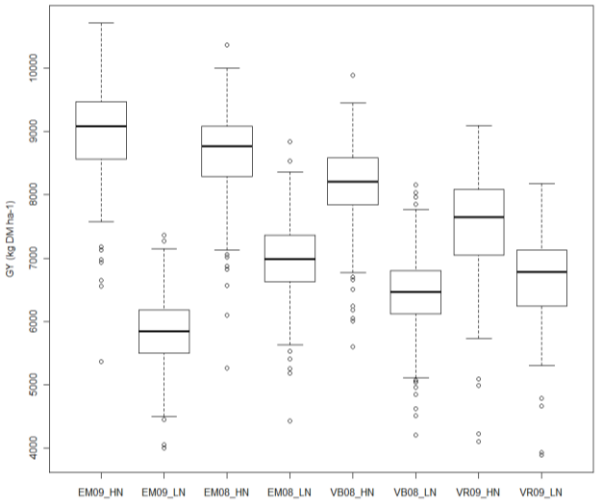


Figure 1: Boxplot of GY for 225 wheat cultivars grown over two years (2008 and 2009) at two N levels [Low N (LN) and High N (HN) and in three sites, Estrées-Mons (EM), Villiers-le-Bâcle (VB) and Vraux (VR)]. Quartiles and median are used to construct the box. The whiskers extend to 1.5 times the interquartile range from the box.

Mean grain yield ranged from 5.8 in EM09_LN to 9.0 t ha^{-1} in EM09_HN (Fig. 1). In all environments, the N effect was always significant with large differences between sites and seasons. Extreme reductions of 11% in VR09 and 35% in EM09 were observed on yield when plants were grown under LN compare to HN conditions. A high correlation between GY measured at HN and LN exists (r=0.86, P<0.001). Older varieties yielded less than the most recent (Fig. 2) suggesting genetic improvement at both HN and LN. NUE was greater at LN (42.7 kg DM kg^{-1} N) than at HN (32.9 kg DM kg^{-1} N).

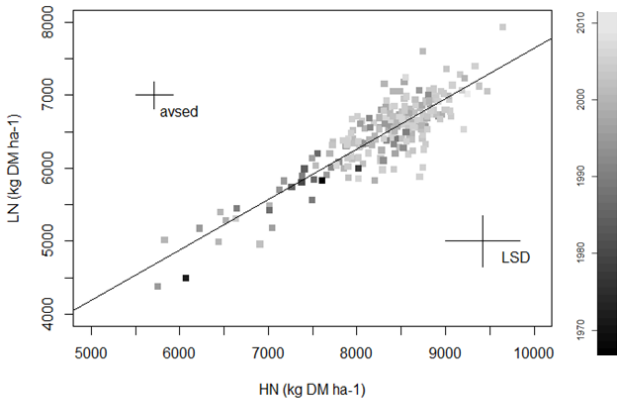


Figure 2: Grain yield best linear unbiased estimators (BLUEs) at low N level (LN) as a function of BLUEs at high N level (HN) for 225 wheat cultivars grown in four environments. Dot colours are function of the year of release from the older (black) to the younger (light grey). Average pairwise prediction standard error (avsed) and least significant difference (LSD) at both HN and LN treatments are plotted as the following regression function: $y = 0.69x + 458.5$ ($r^2 = 0.74$, $P < 0.001$).

NutE was higher at LN (55.6 kg DM kg^{-1} N) than at HN (41.9 kg DM kg^{-1} N), while NupE remained stable (79 % at HN and 78 % at LN). Phenotypic correlations revealed that the contribution to NUE of N utilisation increased with N supply, from r = 0.53 (P < 0.001) at LN to r = 0.60 (P < 0.001) at HN. The contribution of N uptake to NUE is also significant (r = 0.44, P < 0.001) but did not vary between LN and HN.

Variance components and heritability

Significant genotypic effects were observed for all traits except NTA and NupE (Table 2). Trait heritabilities were highly variable ranging from 0 for NupE to 0.97 for flowering date. The high

contribution of the G × E interaction to the genetic variance of N uptake (77 % of the total variance) is consistent with a weak genetic additive effect. HI, NutE, GPD, NutE_Prot, NUE, and NUE_Prot, are all derived traits which nevertheless exhibited high heritabilities.

The variance decomposition revealed significant G × N interactions for GY, GPC, NUE, NUE_Prot, and %N_S. G × N interaction was the most important for NutE representing 7 % of its genetic variance. We should stress that genotype × environment × N interaction was included in the model residual, resulting in an underestimation of the specific influence of N treatment on genotypes.

Heritabilities at HN and LN were really similar (Suppl. data 4). The highest difference was observed for GNY with heritability 0.31 at HN and 0.19 at LN. Nevertheless, differences in variance components should be noticed. For DMGY, GPC, GPD, SA, TKW, NHI, %N_S, and ADM_S genetic and error variances decreased from HN to LN. On the contrary, traits associated with NUE (NutE, NutE_Prot, NUE, and NUE_Prot) have genetic and error variances increasing from HN to LN.

Table 2: Mean, standard deviation (sd), heritability (h^2g) and genetic variance decomposition for agronomic traits measured on 225 wheat cultivars in eight trials (see text for traits description). Genetic variances are decomposed into three components, G the additive genetic effect, the G × E and the G × N interactions.

Trait	Mean	sd	Units	h^2g	G		G × E		G × N	
FLO	149.25	7.12	days	0.97	92%	***	8%	***	0%	ns.
PH	76.60	8.43	cm	0.89	80%	***	19%	***	0%	ns.
SA	411.97	78.8	nb spike m^{-2}	0.75	69%	***	23%	***	8%	**
TKW	42.45	4.11	g	0.91	83%	***	16%	***	1%	ns.
KS	42.78	8.88	nb kernel per spike	0.77	68%	***	30%	***	2%	ns.
GPC	9.93	2.05	% prot	0.85	71%	***	27%	***	2%	*
GY	7400	1258	kg DM ha^{-1}	0.79	60%	***	36%	***	5%	***
GNY	127.94	35.44	kg N ha^{-1}	0.18	18%	**	74%	***	8%	ns.
GPD	0.00	0.78	% prot	0.71	61%	***	36%	***	3%	ns.
%N_S	0.42	0.13	% N	0.66	56%	***	35%	***	9%	*
ADM_S	7288	1861	kg DM ha^{-1}	0.79	81%	***	18%	***	1%	ns.
HI	50.42	5.67	% DM	0.79	67%	***	32%	***	1%	ns.
NHI	81.15	5.71	% N	0.45	38%	***	55%	***	7%	ns.
NTA	158.46	45.03	kg N ha^{-1}	0.04	16%	ns.	75%	***	9%	ns.
NupE	0.78	0.08	% N	0.00	10%	ns.	77%	***	13%	ns.
NutE	48.80	11.19	kg DM kg^{-1} N	0.79	63%	***	30%	***	7%	***
NutE_Prot	0.07	0.01	% prot kg^{-1}N ha^{-1}	0.83	74%	***	23%	***	3%	ns.
NUE_Prot	0.05	0.0083	% prot kg^{-1}N ha^{-1}	0.83	69%	***	27%	***	4%	***
NUE	37.8	7.69	kg DM kg^{-1} N	0.80	69%	***	26%	***	5%	*

*LTR tests : *** , P-value <0.001 ; **, P-value <0.01; * , P-value <0.05; and ns., non-significant P-value>0.05*

Year of registration effect on genetic additive effect

The effect of year of registration (YR) was tested on the different traits. Additive genetic effects were estimated at both HN and LN. YR effect was either tested alone or taking into account precocity and/or plant height and/or quality classes as covariates. These covariates were themselves first tested for association with YR. Quality classes were not totally homogeneously allocated among years (LSD test P = 0.05, Suppl. data 5). "Very high quality" varieties which have higher GPC (LSD test P = 0.05, Suppl. data 5) were on average significantly older (1999) than "high" and "good quality" varieties (2003). Flowering date was correlated to YR with new cultivars later flowering (+0.18 day year^{-1}). YR had no significant effect on plant height but variation in plant size exists (coefficient of variation = 11 %). The addition of covariates enhanced the accuracy of the genetic progress estimation (Fig. 3). Indeed, sampling bias and miss-adaptation of phenology to the tested environments were corrected.

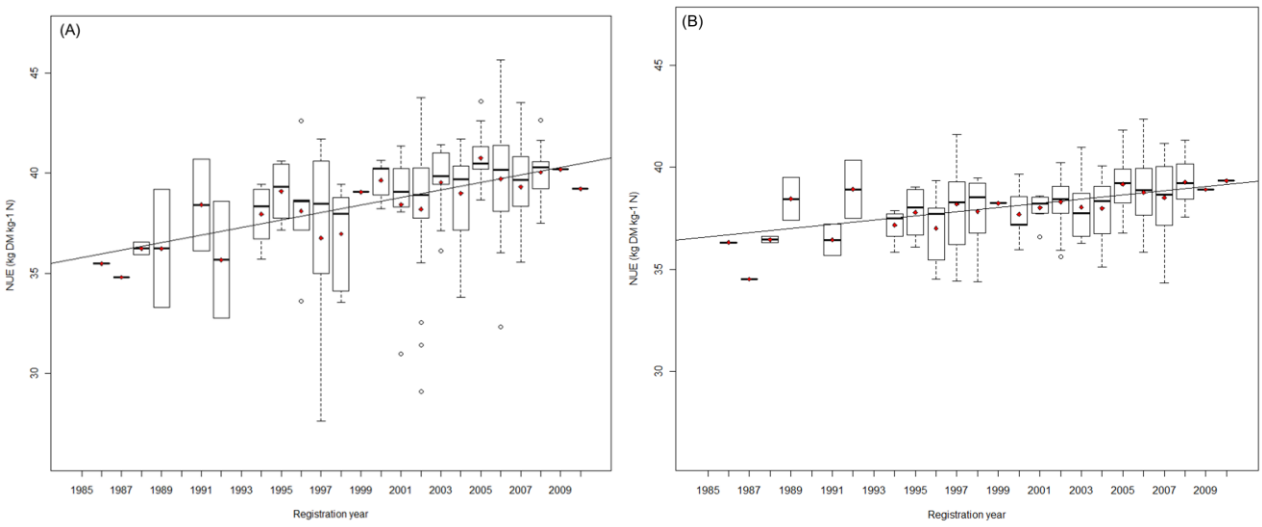


Figure 3: Boxplot of (A) NUE genetic value and (B) NUE genetic values corrected for quality and precocity effects as a function of registration year of 195 wheat cultivars grown in four environments and two N treatments. Medians (dash), means (solid diamond). (A) NUE = 37.8 + (YR - 2002) × 0.198 (r^2 = 12.6 %; P < 0.001). NUE additive genetic values are BLUEs from the multi-environment mixed model. (B) NUE = 37.8 + (YR - 2002) × 0.126; NUE additive genetic values are BLUEs from multi environment mixed model which were corrected for quality and precocity effects. The complete model (with quality, precocity and YR) adjusted r-squared is 64.6 %.

The most significant effect of YR was detected on GY (+0.45 % year^{-1}). GY can be divided into three components: the weight of grains (TKW), the number of grains per spike (KS), and the number of spike per area (SA). TKW and SA remained stable. KS increase was not significant when quality and precocity were added to the model. We can conclude that there is no clear trend about how GY genetic gain was achieved. Probably different strategies have been used simultaneously.

Apart from the variability of quality classes among years, GPC did not decrease since 1985. This stability, coupled with the GY increase, led to GNY improvement (+0.35 % year^{-1}). GNY improvement

can be the result of two physiological changes: partitioning and/or uptake. The YR effect on uptake was not tested as no additive genetic effect was detected for NTA (Table 2). Regarding dry matter partitioning, HI increased (+0.13 % year^{-1}) as ADM_S remained the same and GY increased. Regarding N partitioning, NHI (+0.12 % year^{-1}) increased, ADM_S remained the same and %N_S decreased.The additive genetic effect of NUE increased (+0.33 % year^{-1}) (Fig. 3) thanks to an improvement of NutE (+0.20 % year^{-1}). NutE improvement and NutE_Prot decrease (-0.27 % year^{-1}) revealed that selection has favoured varieties which preferentially convert remobilised nitrogen into grain dry matter rather than into protein. As GPC was stable, the decrease in NutE_Prot (GPC / NTA) could be the result of either NTA improvement or/and an uptake increase. These hypotheses could not be distinguished as no significant additive genetic effect was detected for NupE (Table 2).

Table 3: Year of registration (YR) effects on agronomic traits measured on 195 wheat cultivars grown in eight trials (see text for traits description). YR effect was tested with and without covariates (quality class, precocity, and plant height): contribution to the variance (R^2), factor effect significance (P), and slope of the YR regression (% of the trait mean).

Trait	Only YR			With cofactor and covariates									
				Quality		Precocity		Height		Year of Registration			
	R^2	P	Slope	R^2	P	R^2	P	R^2	P	R^2	P	Slope	
PH	1	ns.		16	***	7	***			0	ns.		
FLO	0	ns.		9	***			7	***	3	**	0.18 day	+0.12%
SA	0	ns.		NT		NT		NT		0	ns.		
TKW	0	ns.		NT		NT		3	**	1	ns.		
KS	2	*	+0.41%	13	***	5	**	NT		0	ns.		
GPC	5	**	-0.46%	52	***	16	***	NT		0	ns.		
GY	17	***	+0.70%	54	***	11	***	NT		6	***	33.2kg DM ha^{-1}	+0.45%
GNY	8	***	+0.38%	5	*	NT		NT		6	***	0.442kg N ha^{-1}	+0.35%
GPD	0	ns.		29	***	5	***	NT		1	ns.		
%N_S	2	*	-0.41%	NT		19	***	12	***	3	**	-2.17×10^{-3} % N	-0.52%
ADM_S	0	ns.		6	***	32	***	16	***	1	ns.		
HI	9	***	+0.29%	41	***	1	*	14	***	2	**	6.71×10^{-2} % DM	+0.13%
NHI	7	***	+0.12%	NT		NT		NT		7	***	9.72×10^{-2} % N	+0.12%
NutE	8	***	+0.39%	40	***	16	***	NT		2	**	9.67×10^{-2} kg DM kg^{-1} N	+0.20%
NutE_Prot	9	***	-0.49%	59	***	10	***	NT		2	***	-1.73×10^{-4} % prot kg^{-1}N ha^{-1}	-0.27%
NUE_Prot	6	***	-0.38%	52	***	12	***	NT		1	ns.		
NUE	13	***	+0.52%	48	***	13	***	NT		5	***	0.13kg DM kg^{-1} N	+0.33%

*Fischer tests : *** , P-value < 0.001 ; **, P-value < 0.01; * , P-value < 0.05; and ns., non-significant P-value > 0.05*

NT = not tested because not significant

YR effect on G × N interaction

After being tested on additive genetic effect, YR effect was tested on significant G×N interactions. A change in G×N interactions was significant only for GY and NUE (Table 4). For GY, the YR effect was significant when no covariates were used. Modern varieties had G×N interaction which increased yield (+0.12 % $year^{-1}$) in HN environments, with a corresponding decrease in LN environments. However, G×N interactions for GY were explained by variation in quality classes (r^2 = 13.1 %, P < 0.001) and precocity (r^2 = 9.8 %, P < 0.001). The most important effect was due to the highly negative interactions of "very high quality" varieties at HN (-188 kg ha^{-1}). The effect of precocity was the result of the positive correlation between date of flowering and G × N interactions at HN (+10 kg ha^{-1} per day of delay). So, once quality and precocity effects were removed, there was no significant difference in GY genetic progress between HN and LN environments (Table 4). The slopes of regression are different but confidence intervals overlap (Fig 4). This also means that recent and old varieties have the same yield loss between HN and LN. However, as recent varieties have a higher GY (+0.35 % $year^{-1}$, Table 3) their relative GY losses are lower than for older varieties and, therefore, recent varieties are more stable.

Concerning NUE, the YR effect on G × N interaction stayed significant when quality was introduced into the model (Table 4). Recent varieties had higher G × N interactions on NUE than older varieties at LN (+2.98 × 10^{-2} kg DM kg^{-1} N $year^{-1}$; +0.08 % $year^{-1}$), and so lower at HN (-2.98 × 10^{-2} kg DM kg^{-1} N $year^{-1}$; -0.08 % $year^{-1}$). The complete genetic progress at LN is calculated as the genetic progress on additive values added to the ones on the G×N interactions. Then, the global genetic progress on NUE was +0.155 kg DM kg^{-1} N $year^{-1}$ at LN and +0.096 kg DM kg^{-1} N $year^{-1}$ at HN (respectively +0.37 % $year^{-1}$ and +0.30 % $year^{-1}$ referring to the mean NUE at LN and at HN) (Fig. 5). This conclusion is consistent with the previous one on GY. Indeed, GY progress was the same at LN and HN; however, N available at LN (mean NTA_{max} = 146.25 kg N ha^{-1}) was lower than at HN (mean NTA_{max} = 231.25 kg N ha^{-1}). So, the way in which NUE is calculated (GY / NTA_{max}) leads to a higher estimate of genetic progress at low N than at high N.

Table 4: Decomposition of G × N interaction variance (%) for NUE and GY of 195 wheat cultivars grown in four environments. The registration year (YR) effect was tested with and without covariates (quality class, precocity, and plant height).

Trait	Only YR	With cofactor and covariates			
		Quality	Precocity	Height	YR
GY	4.6 **	13.09 ***	9.84 ***	NT	1.09 ns.
NUE	3.25 *	5.27*	NT	NT	1.97*

*Fischer tests: ***, P-value < 0.001; **, P-value < 0.01; *, P-value < 0.05 and ns., non-significant P-value > 0.05*

NT = not tested because not significant

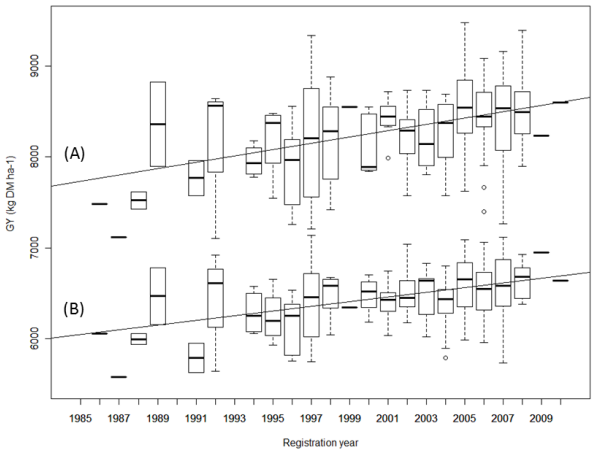


Figure 4: Boxplot of GY genetic values by year of release and by N treatment (LN = low N level; HN = high N level) for 195 wheat cultivars grown in four environments. Values are the best linear unbiased estimators of NUE corrected of quality and precocity effects. (A) at HN treatment, and (B) at LN treatment. (A) At HN, regression function is NUE = - 69690 + YR × (34.8 +/- 4.42), the complete model (with quality and precocity) adjusted r-squared is 66 % and YR effect P<0.001. (B) At LN, regression function is NUE = - 51302 + YR × (25.64 +/- 6.22), the complete model (with quality and precocity) adjusted r-squared is 70 % and YR effect P < 0.001. G × N on NUE are significant but YR effect on this interaction is not significant (P > 0.05).

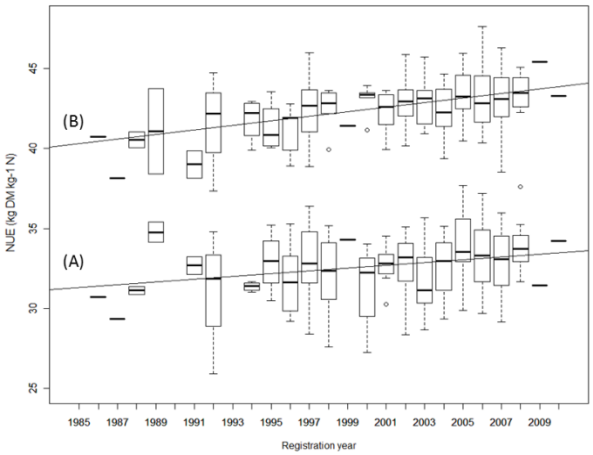


Figure 5: Boxplot of NUE genetic values by year of release and by N treatment (LN = low N level; HN = high N level) for 195 wheat cultivars grown in four environments. Values are the best linear unbiased estimators of NUE corrected of quality and precocity effects. (A) at HN treatment, and (B) at LN treatment. (A) At HN, regression function is NUE = -141.80 + YR × (0.09 +/- 0.03), the complete model (with quality and precocity) adjusted r-squared is 48.8 % and YR effect P < 0.001. (B) At LN, regression function is NUE = - 240.84 + YR × (0.14 +/- 0.02), the complete model (with quality and precocity) adjusted r-squared is 66.2 % and YR effect P < 0.001. G × N on NUE are significant and YR effect on this interaction is significant (P < 0.05).

DISCUSSION

We studied the variance components of NUE among 225 European winter wheat varieties evaluated in 8 independent trials containing two N treatments. These varieties were mostly released between 1985 and 2010. Thus, a study of the genetic improvement of NUE over the past 25 years was possible. We found that using quality, precocity, and plant height, more accurate estimations of genetic gains were possible. The effect of selection was assessed on the additive genetic value and on the G × N interaction term. No additive genetic effect was found on NupE. The high heritability of complex traits such as NutE, NHI, NUE, NutE_Prot, and NUE_Prot revealed their potential for breeding. Regarding additive genetic value, NUE has increased thanks to a rise in NutE. Protein concentration did not decrease since 1985. The main factor in this progress was better partitioning as revealed by an increase in NHI linked to a decrease in straw N concentration at maturity. G × N interactions were significant on GY, NUE, NutE, GPC, and NUE_Prot. Significant changes for G × N interactions were only detected for NUE, attesting to the higher yield stability of recently released compared to older varieties.

Genetic progress assessment method

This work has been carried out with a large collection of European elite winter varieties, which have been bred for different target environments. They were mainly varieties designed for the French market and also for neighbouring countries (e.g. Germany, Great Britain, and Italy). In contrast to previous studies on NUE the period under study was smaller and encompasses the last 25 years of breeding, compared to 82 years (Uzik and Zofajova 2012) and 94 years (Guarda et al. 2004) for winter wheat, 35 years (Ortiz-Monasterio et al. 1997a) and 99 years (Muurinen et al. 2006) for spring wheat, and 75 years for barley (Bingham et al. 2012). Therefore, the period under study did not include major selection events that took place for plant height and precocity in the previous periods. It turns out that, in our panel, mean height was 78.9 +/- 8 cm at HN and was stable over years (Table 3). This value is very similar to the 80 cm reported by Gooding et al. (2012) as being optimum for NUE using near isogenic lines for different *Rht-1* alleles. Nevertheless, variability existed in our panel (Table 2), and had to be controlled to avoid interference in breeding effect estimation.

Precocity was also controlled by flowering date assessment. In our panel, the delay in flowering date is explained by the non-homogenous distribution of the varieties' origins (Suppl. data 6). Varieties bred to European northern countries are generally late (Worland 1996) and are more frequent among the recent varieties of our panel. After 2005, four varieties came from the south of Europe (Italy, Spain) and 10 from the north (*e.g.* Germany, Great-Britain, Denmark).

In the same way, we chose to control for quality class. Two points have to be addressed. First, "very high quality" varieties are often high GPC varieties. A negative correlation between GY and GPC exists (*e.g.* Simmonds 1995; Oury and Godin 2007; Bogard et al. 2010), and so NUE and GPC are negatively correlated (Barraclough et al. 2010; Gaju et al. 2011). These low-yielding genotypes can bias the analyses if they are not evenly distributed over time. Secondly, Ortiz-Monasterio et al. (1997b) studied genetic progress for grain quality from 1950 to 1985, and found no link between quality (alveograph's parameters) and YR. Guarda et al. (2004) also studied wheat quality evolution between 1900 and 1994. They concluded that lower protein concentration was associated with an improvement in protein composition, resulting in an increase of bread-making quality. Moreover, "very high quality" varieties frequency does not drastically vary among years, according to the French official catalogue of registered bread wheat varieties. So, in our case having older "very high quality" varieties was a sampling bias (Suppl. data 5) that had to be controlled.

As with other field studies on NUE genetic progress, we did not take into account below-ground dry matter. However, not taking into account roots in the determination of N related traits such as NupE appears of little influence (Allard et al. 2013). Significant genotypic differences for root N exist but the amount of N present is low compared to total plant N. And so, genotype ranking is not affected.

Genetic progress between 1985 and 2010

This study concludes that significant grain yield (GY) improvement is observed at both HN and LN. The genetic gain on GY is estimated to be +0.45 % $year^{-1}$ (+33.2 kg DM ha^{-1} $year^{-1}$) with no significant difference between HN and LN. This linear trend is in agreement with the requirement that a variety has to yield in excess of control varieties in official trials to be registered in France. The control variety list evolves to be representative of their market shares and agricultural practices. Progress on GY was not related to progress on TKW, SA, or KS. This is in contrast with Brancourt-Hulmel et al. (2003) who studied GY evolution by comparing 14 winter wheat cultivars registered between 1946 and 1992 in France at two levels of fungicide and N treatments and concluded that GY improvement was made by an increase in kernel number. Our study suggests a diversification of strategies in a more recent period.

Concerning differences between HN and LN treatment, Ortiz-Monasterio et al. (1997a), Brancourt-Hulmel et al. (2003), Guarda et al. (2004) concluded that GY progress was higher at HN than at LN. But these studies were based on mean differences in N treatment and not on G × N coefficients. Also, according to Ortiz-Monasterio et al. (1997a), this difference was not significant for the period 1962 to 1985. Moreover, in Brancourt-Hulmel et al. (2003) and Guarda et al. (2004), no fertilisers were added in the very low N treatment. In contrast, in this study, varieties' behaviours were assessed in a context of fertiliser reduction.

We also showed that grain protein concentration (GPC) did not significantly change in the last 25 years. At HN, the mean GPC of bread-making wheat ("very high quality", "high quality", and "good quality") was 11.4 +/- 1.6 %. This content is sufficient to fulfil French milling demands and exportation requirements to North Africa, the main exportation area for French production. Selection on GPC may only result in the elimination of low GPC lines and not in increasing GPC. Breeding program objectives were clearly to increase GY and maintain quality. But, in this study, mean GPC at LN is 8.66 +/- 1.62 % which is largely below bread-making and exportation requirements. If suboptimal conditions are targeted, one of the main challenges for breeders will be to considerably increase GPC. An alternative would be to modify protein composition to increase dough strength and viscoelasticity, allowing for lower protein grain to be suitable for bread-making.

Brancourt-Hulmel et al. (2003) assessed a genetic gain of +0.15 % $year^{-1}$ for NHI between 1946 and 1992, which includes semi-dwarf allele integration in breeding programs, compared to +0.12 % $year^{-1}$ in our study. These two estimates are very similar. An explanation is that there is no statistically significant increase in NHI from adding single semi-dwarf alleles to a tall background (Gooding et al. 2012). Besides, the absence of a link between quality and NHI is confirmed by Barraclough et al. (2010) who compared 39 elite commercial cultivars during four years at five N rates. This suggests an equivalent N partitioning between varieties from different quality classes. N absorbed before flowering, stored in vegetative parts and then remobilised to the grain accounts for around 70 % of total grain N (Van Sanford and MacKown 1986; Kichey et al. 2007). We found that the NHI increase was associated with a %N_S decrease (and ADM_S stability). This better N partitioning may either

come from a more efficient N remobilisation and/or a more efficient translocation efficiency (N absorbed after anthesis and translocated to the grain, Kichey et al. 2007).

Nitrogen use efficiency improvement was mainly due to better N utilisation efficiency. Our estimations of genetic progress were in the range of previously published results, even if the N available was estimated differently. This study assessed NUE genetic progress of +0.37 % year^{-1} at LN and +0.30 % year^{-1} at HN. Ortiz-Monasterio et al. (1997a) reported that NUE genetic progress was 0.4-1.1 % year^{-1} depending of N applied for spring CIMMYT cultivars released between 1950 and 1985. Sylvester-Bradley and Kindred (2009) also reported a significant trend between old and new cultivars grown at 0 and 200 kg N ha^{-1}. In contrast, Muurinen (2006) concluded a lack of genetic gain on NUE for 18 spring wheat varieties bred between 1901 and 2000. As in our study, various reports have shown a major effect of N utilisation compared to N uptake on NUE at high N input (Ortiz-Monasterio et al. 1997a; Brancourt-Hulmel et al. 2003; Uzik and Zofajova, 2012). In contrast, at low N input, N uptake seems to be the component which has more effect on NUE (Ortiz-Monasterio et al. 1997a; Le Gouis et al. 2000; Muurinen et al. 2006). In our study, NupE contribution to NUE was the same at LN and HN treatments, and the additive genetic effect on NupE was not significant. So, detection of change on NupE was impossible.

To better compare the different studies, a finer characterisation of the N status at different N levels is probably necessary. In their low N input level, Ortiz-Monasterio (1997a), Le Gouis (2000), and Muurinen (2006) added no N fertiliser. Only mineral N already present in the soil and N coming from the mineralisation of organic matter were available to the plants. Our LN input modality was less stressful with a mean of 130 kg ha^{-1} (fertiliser + soil N) available to crop.

Three hypotheses can account for the absence of an additive genetic effect of NupE in this study. (i) Genetic variation on uptake may only appear in highly N deficient environments. Indeed, NupE genetic variances are very similar between HN and LN (Suppl. data 4). But this hypothesis contradicts the single trial analysis (data not shown) where NTA genetic additive effect was significant only in two HN trials (VR09_HN and EM08_HN). (ii) The common method of using pre-sowing or post-winter early measurements of soil mineral N clearly underestimates NupE, as N losses (e.g. leaching, volatilisation) are not taken into account and so available N is overestimated. At the opposite extreme, the risk of overestimating NupE is real at LN as mineralization can provide N in large quantities and leaching is limited so that available N is underestimated. For example, Ortiz-Monasterio (1997a), Le Gouis (2000), and Muurinen (2006) used this method and reported NupE superior to 1 in their low N input trials. Bingham et al. (2012) showed that the method of calculation had little effect on relative differences between varieties in single N treatment analysis as NTA between methods are only divided by different coefficients to obtain NupE. But when different N levels are used in common analysis, if overestimation bias at LN is not compensated by the underestimation bias at HN, this can lead to misinterpretation. To avoid this, we chose here (and advocate) to use the maximal uptake measured at each N level. To take into account possible measurements errors we used the 95th percentile.

(iii) The genetic variation of uptake is not sufficient in our panel in comparison to the precision of measurements included in the computation of NTA / NupE. Measurement errors could be controlled using more replicates or larger sampling size but with an additional cost. In addition, variability may have to be researched in a more diverse panel using for example genetic resources or breeding materials.

Breeding efficiencies for different N levels

Falconer and Mackay (1996) formulated that the relative efficiency under direct selection in condition 1 versus indirect selection in condition 2 is $r_{G12} \times h_2 / h_{1,}$ where h_1 and h_2 are heritabilities in the two conditions respectively and r_{G12} the genetic correlation between conditions. Heritability is usually lower under LN conditions (Brancourt-Hulmel et al. 2005, Laperche et al. 2006a), suggesting that indirect selection at high N can be an effective strategy to breed for low N conditions. In maize, Presterl et al. (2003) advocated direct selection at LN when yield reduction is > 21 % based on the evolution of the genetic correlation as a function of yield reduction. For Anbessa et al. (2010) indirect selection was efficient in barley, but the estimation was made on data where yield reduction was only 7 %. In a study where yield was reduced on average by 35 %, Brancourt-Hulmel (2005) advised to directly select wheat in LN environments to maximise gains. In this study, the mean yield in LN trials was reduced by around 20% compared to the mean yield in HN trials.

Genetic progress on NUE and NUE-related traits was assessed from the additive genetic effect estimated using both HN and LN levels together with the G × N interaction. Our work shows that recent varieties have enhanced NUE-associated traits at both LN and HN treatments (except in N utilisation for protein, NutE_Prot). The only significant genetic progress difference occurred for NUE; +0.37 % $year^{-1}$ and +0.30 % $year^{-1}$ respectively at LN and HN. The varieties we used were probably mostly selected in HN environments as usually done in private breeding programs. Using the formula cited above, we calculated that the relative efficiency for indirect selection at HN for LN conditions was 78.1 % for NUE. This was mostly due to the fact that heritabilities were similar in our conditions at LN and HN. We advise to directly select in N suboptimal conditions when moderate N stressful environments are targeted.

Around 10 years are needed for making crosses giving thousands of progenies to register a new variety. As the number of selected lines is reduced, the range of environments in which they are tested is wider. Among all these trials, moderate N stresses surely occur. So the selection process may already mixes HN and moderate LN environments explaining in part the similar genetic progress at HN and at LN. Nevertheless, this selection regime has to be consciously designed to make it more efficient. We can imagine characterizing the N constraint using control varieties repeated in each trial for which NTA will be calculated, measuring %N_S and ADM_S. Selection will then be made only using trials where the chosen stress effectively occurred.

NUE enhancement actually arises from selection on yield. Indeed, screening for NUE components is time consuming and may not be implemented in breeding programs soon. High-throughput methods are currently being developed (Tester and

Langridge 2010) but are not yet adapted to the thousands of lines that are tested in a breeding program. Therefore, improvement focused on NutE or NupE will be conditioned by the possibility to perform molecular selection on QTLs or genes. A few studies have already identified chromosomal regions associated with these traits using wheat plants grown in the field or in controlled conditions (*e.g.* Laperche et al. 2006b; Bordes et al. 2012; Guo et al. 2012, Liu et al. 2013). Understanding root architecture and its interaction with N supply is also one promising way to improve NUE in plants (Hirel et al. 2007; Foulkes et al. 2009; Kant et al. 2011). But phenotyping of wheat roots in the field is complex (for a review see Manske et al. 2001). As high throughput screens in the field are not available yet, genetic progress will also depend on the development and the use of molecular markers for enhanced root systems.

Root architecture is also affected by the *Rht* dwarfing genes (Laperche et al. 2006b; Wojciechowski et al. 2009) which were the main factors of wheat improvement in the world. Dwarfing alleles are widely spread and used to control response to high N supply by reducing response to gibberellin acid (GA) and thus plant height (Peng et al. 1999) and lodging (Ortiz-Monasterio et al. 1997a). Laperche et al. (2006b) reported a negative effect of dwarfing alleles on both root and aerial biomass of young plants grown at low N in controlled conditions. In this study, varieties have different dwarfing genes to achieve short height. Moreover, frequencies of the combination of the GA-insensitive dwarfing alleles (*Rht-B1* and *Rht-D1*) changed as a function of the year of registration (HSD test P = 0.05; Suppl. data 7A). When dwarfing allele combinations were used in the model of genetic progress assessment, it appeared that they explained more of the G × N variance to NUE than YR. But they had no effect on NUE additive genetic values (Suppl. Table 7B). Recent varieties have G × N interactions which enhanced their NUE at LN, and so may have a more stable yield also because of the introduction of *Rht-D1b*. In contrast, this stability in yield also means that recent varieties are capitalised less on N input increase than older ones. This may be a consequence of GA-insensitivity as GA has a major role in regulating developmental processes (Hedden, 2003). So, the use of alternate GA-sensitive dwarfing alleles such as *Rht8c* needs to be tested. Indeed Gooding et al. (2012) studied near isogenic lines and concluded that at anthesis the *Rht8c* + *Ppd-D1a* (dwarf and photo-insensitive) line accumulated similar quantities of nitrogen to *Rht-D1b* despite its earliness (due to its photoperiod-insensitivity).

CONCLUSIONS

In a global context of fertiliser reduction, we investigated nitrogen use efficiency improvement using a European panel of elite winter wheat cultivars. This study is one of the first to use so many varieties in a multi-environment direct comparison between old and recent varieties. Quality, precocity, and height were used to control panel heterogeneity. Variance decompositions were used to describe the genetic determinism of NUE-related traits and to identify significant G × N interactions. We report equal genetic progress at both HN and LN treatments for all traits except for NUE, which were significantly enhanced at both N

levels but more efficiently at LN. This demonstrates the higher yield stability of recent varieties. We conclude that direct selection in HN conditions for LN conditions is efficient, but advise to directly select at LN if this is the targeted treatment. Two major challenges now appear. The first challenge will be to increase GPC at LN; and the second will be to increase uptake efficiency while maintaining utilisation efficiency improvement.

ACKNOWLEDGEMENTS

Data were obtained thanks to the support of the ANR ProtNBle project (06 GPLA016). The authors would like to thank the staff at Estrées-Mons experimental unit (INRA) and at Villiers-le-Bâcle and Vraux experimental farms (Arvalis) that have phenotyped so many traits. The authors are also grateful to the ANRT (Association Nationale de la Recherche et de la Technologie) which support the PhD thesis (129/2012) during which the analyses were made. Sincere thanks to Ian Mackay for his editorial advice.

AUTHOR CONTRIBUTION STATEMENT

K Beauchêne and E Heumez were respectively in charge of field trials conducted by Arvalis and INRA. Statistical analyses were conducted by F Cormier during his PhD thesis co-directed by S Praud and J Le Gouis. F Cormier and J Le Gouis wrote the manuscript. S Faure, P Dubreuil, S Lafarge, and S Praud provided useful help on methods, interpretations, and in reviewing the manuscript.

REFERENCES

Anbessa Y, Juskiw P, Good A, Nyachiro J, Helm J (2010) Selection efficiency across environments in improvement of barley yield for moderately low nitrogen environments. Crop Sci 50:451-457

Allard V, Martre P, Le Gouis J (2013) Genetic variability in biomass allocation to roots in wheat is mainly related to crop tillering dynamics and nitrogen status. Eur J Agron. 46:68-73

Atlin GN, Frey KJ (1989) Predicting the relative effectiveness of direct versus indirect selection for oat yield in three types of stress environments. Euphytica 44:137-142

Austin RB (1999) Yield of wheat in the United Kingdom: recent advances and prospects. Crop Sci 39:1604-1610

Bänziger M, Etran FJB, Afitte HRL (1997) Efficiency of high-nitrogen selection environments for improving maize for low-nitrogen target environments. Crop Sci 37:1103-1109

Barraclough PB, Howarth JR, Jones J, Lopez-Bellido R, Parmar S, Shepherd CE, Hawkesford MJ (2010) Nitrogen efficiency of wheat: genotypic and environmental variation and prospects for improvement. Eur J Agron 33:1–11

Bingham I, Karley A, White P, Thomas W, Russell J (2012) Analysis of improvements in nitrogen use efficiency associated with 75 years of spring barley breeding. Eur J Agron 42:49-58

Brancourt-Hulmel, M, Doussinaut G, Lecomte C, Berard P, LeBuanec B, Trottet M (2003) Genetics improvement of agronomic traits of winter wheat cultivars released in France from 1946 to 1992. Crop Sci, 43: 37-45

Bogard M, Allard V, Brancourt-Hulmel M, Heumez E, Machet JM, Jeuffroy MH, Gate P, Martre P, Le Gouis J (2010) Deviation from the grain protein concentration-grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. J Exp Bot 61:4303-4312

Bordes J, Ravel C, Jaubertie JP, Duperrier B, Gardet O, Heumez E, Pissavy AL, Charmet G, Le Gouis J, Balfourrier F (2012) Genomic regions associated with the nitrogen limitation response revealed in a global wheat core collection. Theor Appl Genet 126:805-822

Brancourt-Hulmel M , Heumez E, Pluchard P, Beghin D, Depatureaux C, Giraud A, Le Gouis J (2005) Indirect versus direct selection of winter wheat for low input or high input levels. Crop Sci 45:1427-1431

Brisson N, Gate P, Gouache D, Charmet G, Oury F-X, Huard F (2010) Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. Field Crops Res 119:201-212

Bruinsma J (2009) The resource outlook to 2050. By how much do land, water use and crop yields need to increase by 2050? FAO, Expert Meeting on How to Feed the World in 2050, Rome, Italy

Bushuk W (1998). Wheat breeding for end-product use. Euphytica 100:137-145

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) ASReml-R Reference Manual." Queensland Department of Primary Industries. http://www.vsni.co.uk/downloads/asreml/release2/doc/asreml-R.pdf

Ceccarelli S, Grando S, Hamblin J (1992) Relationship between barley grain yield measured in low- and high-yielding environments. Euphytica 64:49-58

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Envir S 11:381-393

Erisman JW, Galloway JA, Sutton MS, Klimont Z, Winiwater W (2008). How a century of ammonia synthesis changed the world. Nat Geosci 1:636-639

Falconer D, Mackay T (1996) Introduction to quantitative genetics. 4th ed. Longman Scientific & Technical, New York

Fischer R, Edmeades G (2010) Breeding and cereal yield progress. Crop Sci 50:85-98

Foulkes M, Hawkesford M, Barraclough P, Holdsworth M, Kerr S, Kightley S, Shewry P (2009) Identifying traits to improve the nitrogen economy of wheat: recent advances and future prospects. Field Crops Res 114:329–342

Gaju O, Allard V, Martre P, Snape JW, Heumez E, Le Gouis J, Moreau D, Bogard M, Griffiths S, Orford S, Hubbart S, Foulkes MJ (2011) Identification of traits to improve the nitrogen-use efficiency of wheat genotypes. Field Crops Res 123:139-152

Gooding M J, Addisu M, Uppal RK, Snape JW, Jones HE (2012) Effect of wheat dwarfing genes on nitrogen-use efficiency. J Agric Sci 150:3-22

Goulding KWT (2004). Minimising losses of nitrogen from UK agriculture. J R Agric Soc Engl 165:1-11

Graybosch R, Peterson C (2012) Specific adaptation and genetic progress for grain yield in Great Plains hard winter wheats from 1987 to 2010. Crop Sci 52:631-643

Green A, Berger G, Griffey C, Pitman R, Thomason W, Balota M, Ahmed A (2012) Genetic yield improvement in soft red winter wheat in the eastern United States from 1919 to 2009. Crop Sci 52:2097-2108

Guarda G, Padovan S, Delogu G (2004) Grain yield, nitrogen-use efficiency and baking quality of old and modern Italian bread-wheat cultivars grown at different nitrogen levels. Eur J Agron 21:181-192.

Guo Y, Kong FM, Xu YF, Zhao Y, Liang X, Wang YY, An DG, Li SS (2012) QTL mapping for seedling traits in wheat under varying concentrations of N, P and K nutrients. Theor Appl Genet 124:851-865

Hedden P (2003). The genes of the green revolution. Trends Genet 19:5-9

Hirel B, Le Gouis J, Ney B, Gallais A (2007) The challenge of improving nitrogen use efficiency in crop plants: toward a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Bot 58:2369-2387

Kant S, Bi YM, Rothstein S (2011) Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. J Exp Bot 62:1499-1509

Kendall MG, Stuart A (1979) The Advanced Theory of Statistics, 4th ed, Griffin, London

Kibite S, Evans LE (1984) Cause of negative correlations between grain yield and grain protein concentration in common wheat. Euphytica 33:801-810

Kichey T, Hirel B, Dubois F, Le Gouis J (2007) In winter wheat (Triticum aestivum L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. Field Crops Res 102:22-32

Laperche A, Brancourt-Hulmel M, Heumez E, Gardet O, Le Gouis J (2006a) Estimation of genetic parameters of a DH wheat population grown at different N stress levels characterized by probe genotypes. Theor Appl Genet 112:797-807

Laperche A, Devienne-Barret F, Maury O, Le Gouis J, Ney B (2006b) A simplified conceptual model of carbon/nitrogen functioning for QTL analysis of wheat adaptation to nitrogen deficiency. Theor Appl Genet 113:1131-1146

Le Gouis J, Beghin B, Heumez E, Pluchard P (2000) Genetic differences for nitrogen uptake and nitrogen utilisation efficiencies in winter wheat. Eur J Agron 12:163-173

Le Gouis J (2011) Genetic improvement of nutrient use efficiency in wheat. In: (ed) The molecular and physiological basis of Nutrient Use Efficiency in Crops. Wiley-Blackwell, Oxford, pp.123-139

Liu X, Li R, Chang X, Jing R (2013) Mapping QTLs for seedling root traits in a doubled haploid wheat population under different water regimes. Euphytica 189:51-66

Lopez-Bellido L, Lopez-Bellido R, Lopez-Bellido F (2006) Fertilizer nitrogen efficiency in durum wheat under rainfed Mediterranean conditions: effect of split application. Agron J 98:55-62

Lopez MS, Reynolds MP, Manes Y, Singh RP, Crossa J, Braun HJ (2012) Genetic yield gains and changes in associated traits of CIMMYT spring bread wheat in a “historic” set representing 30 years of breeding. Crop Sci 52:1123-1131

Manske GGB, Ortiz-Monasterio IJ, Vlek PLG (2001) Techniques for measuring genetic diversity in

roots. In: Moll RH, Kamprath EJ, Jackson WA (1982) Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization. Agron 74:562-564

Monaghan JM, Snape JW, Chojecki AJS, Kettlewell PS (2001) The use of grain protein deviation for identifying wheat cultivars with high protein concentration and yield. Euphytica 122:309-317

Muurinen S, Slafer GA, Peltonen Sainio P (2006) Breeding effects on Nitrogen Use Efficiency of Spring Cereals under northern Conditions. Crop Sci 46:561-568

Olesen JE, Jørgensen LN, Petersen J, Mortensen JV (2003). Effects of rate and timing of nitrogen fertilizer on disease control by fungicides in winter wheat. 2. Crop growth and disease development. J Ag Sci 140:15-29

Oury FX, Bérard P, Brancourt-Hulmel M, Depatureaux C, Doussinault G, Galic N, Giraud A, Heumez E, Lecomte C, Pluchard P, Rolland B, Rousset M, Trottet M (2003) Yield and grain protein concentration in bread wheat: a review and a study of multi-annual data from a French breeding program. J Genet Breed 57:59-68

Oury FX, Godin C (2007) Yield and grain protein concentration in bread wheat: how to use the negative relationship between the two characters to identify favourable genotypes? Euphytica 157:45-57

Oury FX, Godin C, Mailliard A, Chassin A, Gardet O, Giraud A, Heumez E, Morlais JY, Rolland B, Rousset M, Trottet M, Charmet G (2012) A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France. Eur J Agron 40:28-38

Ortiz-Monasterio I, Sayre KD, Rajaram S, McMahon M (1997a) Genetic progress in wheat yield and nitrogen use efficiency under four N rates. Crop Sci 37:898-904

Ortiz-Monasterio I, Pena RJ, Sayre KD, Rajaram S (1997b) CIMMYT's genetic progress in wheat grain quality under four N rates. Crop Sci 37:892-898

Pathak RR, Lochab S, Raghuram N (2011) Plant Systems | Improving Plant Nitrogen-Use Efficiency. In: Murray Moo-Young (ed.), Comprehensive Biotechnology, 2nd edn. Elsevier, pp 209-218

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999)'Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256-261

Presterl T, Seitz G, Landbeck M, Thiemt W, Schmidt W, Geiger HH (2003) Improving nitrogen use efficiency in European maize: estimation of quantitative parameters. Crop Sci 43:1259-1265

Rothstein S (2007) Returning to our roots: making plant biology research relevant to future challenges in agriculture. Plant Cell 19: 2695-2699

Shewry PR, Halford NG (2002). Cereal seed storage proteins: Structures, properties and role in grain utilization. J Exp Bot 53:947-958

Shewry PR (2004) Improving the protein content and quality of temperate cereals: wheat, barley and rye. In Cakmak I, Welch R (ed) Impacts of agriculture on human health and nutrition. USDA, ARS, U.S. Plant, Soil and Nutrition Laboratory, Cornell University, USA

Simmonds NW (1995). The relation between yield and protein in cereal grain. J Sci Food Agric 67:309-315

Sinebo W, Gretzmacher R, Edelbauer A (2002) Environment of selection for grain yield in low fertilizer input barley. Field Crops Res 74:151-162

Sylvester-Bradley R, Kindred DR (2009) Analysing nitrogen responses of cereals to prioritize routes to the improvement of nitrogen use efficiency. J Exp Bot 60:1939-1951

Tester M, Langridge P (2010) Breeding Technologies to Increase Crop Production in a Changing World. Sci 327:818-822

Uzik M, Zofajova A (2012) Progress in accumulation and translocation of nitrogen in winter wheat cultivars released in the years 1921-2003. Cereal research communications 40:135-146.

Van Sanford DA, MacKown CT (1987) Cultivar divergences in nitrogen remobilization during grain filling in soft red winter wheat. Crop Sci 27:295-300

Wojciechowski T, Gooding M, Ramsay L, Gregory PJ (2009) The effects of dwarfing alleles on seedling root growth of wheat. J Exp Bot 60:2565-2573

Worland AJ (1996) The influence of flowering time gene on environmental adaptability in European wheats. Euphytica 89:49-57.

A GWAS-BASED METHOD TO SPEED UP QTL CLONING

Past nitrogen use efficiency (NUE) improvement was mainly driven by selection on grain yield while maintaining grain protein content. Nevertheless, to deal with the fertiliser reduction advocated by political, economic, and environmental concerns, genetic progresses should be accelerated. Due to the difficulties linked to NUE phenotyping methods (partially destructive and laborious); we suggest the use of genetic markers as a promising way to achieve future genetic progresses. In this sense, here, we will discuss about gene discovery using genome-wide association studies (GWAS). This was also the topic of a talk made at the Plant and Animal Genome conference (January 2015, San Diego).

Speeding QTL cloning

The most performant way to screen for varieties based on quantitative trait loci (QTL) is to use genetic markers tagging causal mutations in genes significantly involved in the studied trait. For this purpose, these genes and their polymorphisms should be known. Seeking for locus involved in a trait and refining the genetic/physical distance to be able to identify candidate genes is classically named "QTL cloning".

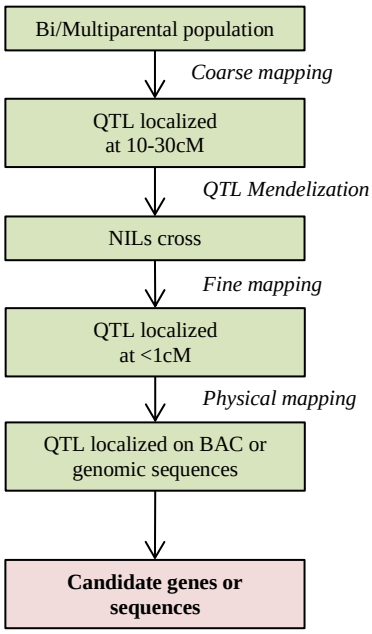


Figure 6: Flow-chart of quantitative traits dissection. (From Salvi and Tuberosa, 2005)

Looking at the flow chart of a classical QTL cloning approach (Fig. 6), we understand that for winter hexaploid wheat this process can be long and fastidious due to the genome complexity and the life cycle

length (1-3 generations per year). In addition, it requires developing a sufficient number of genotypes accumulating a sufficient number of recombination to actually end with a few candidate genes. Moreover, the studied diversity is directly linked to the diversity of the parents used to build the mapping population. These three limitations (development of population, mapping resolution and allelic diversity) can be overcome by GWAS approaches at the cost of the statistical power of detection. Indeed, although smaller linkage disequilibrium (LD) increases mapping resolution, it decreases linkage disequilibrium between causal mutations and genetic markers. Thus, a question arises: "In wheat, could we speed up QTL cloning using GWAS?"

Defining QTL boundaries

The concept of QTL only makes sense if we are able to define locus boundaries. In multiparental design, methods to define boundaries from QTL mapping results are commonly used (*e.g.* LOD support interval, bootstrapping). In GWAS, results are mostly published only as Manhattan plots [-log(P-value) as a function of genomic coordinates] focusing on significant spots (quantitative trait nucleotide, QTN) and not on regions (QTL). Nevertheless, in the few studies aiming to define QTL from QTN information, the use of the mean LD decay appeared to be a consensus method (Tian et al. 2011; Zhao et al. 2011; Le Gouis et al. 2012). But, using the mean LD decay may not be sufficient as LD is highly variable. For example, meiotic recombination rate (a component of LD) fluctuates significantly (Fig. 7). Thus, a more accurate method should be developed.

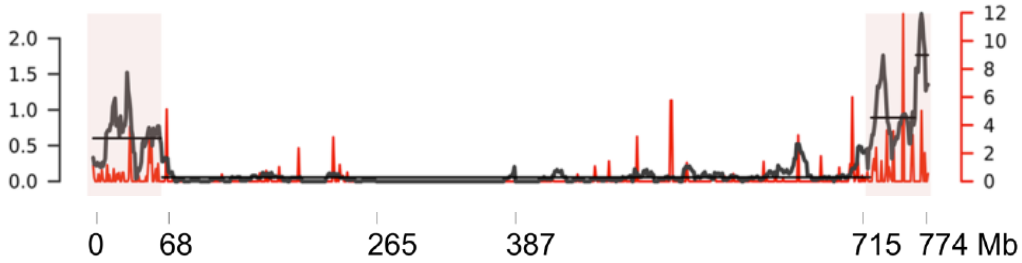


Figure 7: Meiotic recombination rate on wheat chromosome 3B (cM/Mb). Sliding window of 10 Mb in black and 1 Mb in red (Choulet et al. 2014).

What are false positives?

"In GWAS, false positive are a major issue." In fact, this common statement always refers to false positive SNP-trait associations (rejection of the H_0 hypothesis of no marker-trait association while it is true) which can arise from population structure (long distance LD) and multiple testing. To deal with population structure, several models (*e.g.* model Q: groups of structure as a covariate or model K: kinship matrix to modelized varieties variance/covariance) have been proposed and/or combined. And, methods to correct for multiple testing are also commonly used (*e.g.* false discovery rate). But, if we are no more focusing on spot (QTN) but on region (QTL), we need to extend our false positive approach to take into account the method used to define QTL from QTN results. Indeed, the fact that a SNP-trait association is true or false is

not the only fact that matters. What also matters is having the causal gene within the QTL boundaries when QTL cloning is at stake. Consequently, a new kind of false positive appears: "false positive QTL" defined as a QTL which do not contain any causal mutation; no matter if the SNP-trait associations (used to build the QTL) were false or true positives. False positive QTL are the real issue in QTL cloning based on GWAS results. Their proportion among positive QTL (all QTL computed from GWAS results) is the main indicator of the efficiency of GWAS-based QTL cloning methods. Thus, power of QTL cloning GWAS-based methods should be studied regarding the entire process: from QTN detection to QTL definition.

A method to define QTL

We developed an empiric method to define QTL from GWAS results based on local LD (Fig. 8) and assessed its power using simulation study. Details will be provided in the next Part of this manuscript. Here, we wanted to focus on the results that contributed to build our gene discovery strategy.

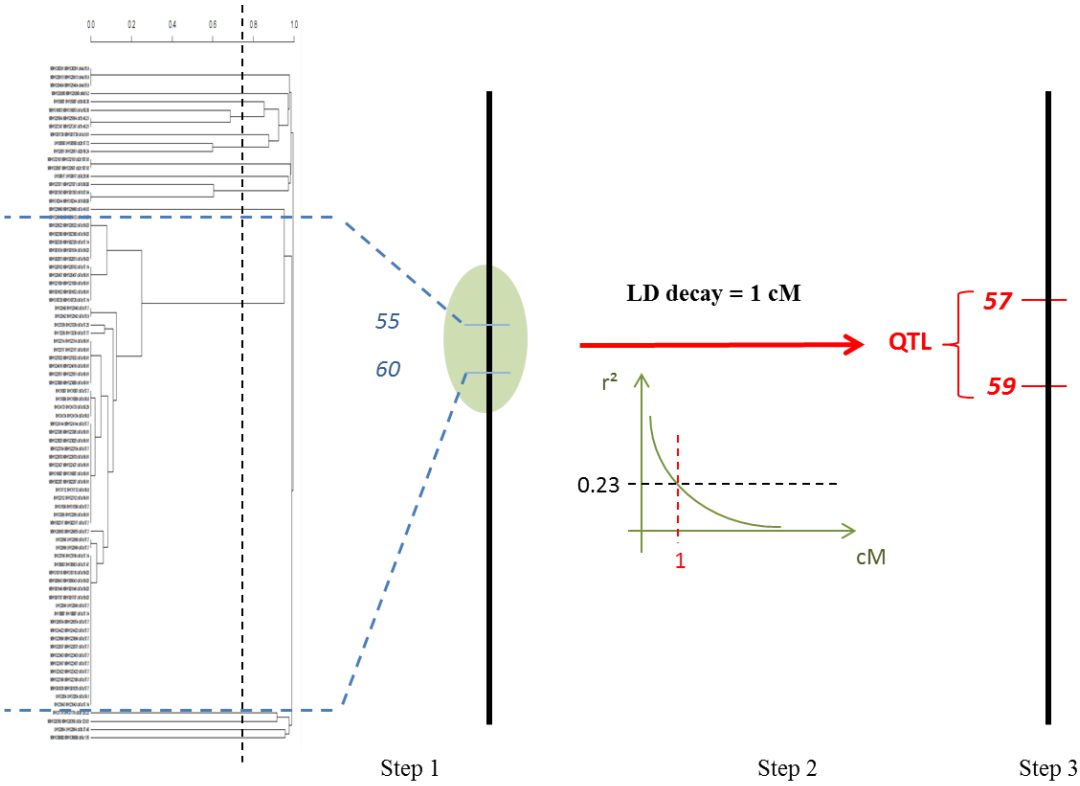


Figure 8: Method used to define QTL from GWAS result. Step 1: QTN clustering in function of LD (r^2) (method average, cut-off = 1- critical LD) to define LD block. Then, QTL first boundaries are defined as the maximum and minimum map positions of QTN belonging to a same LD block. **Step 2:** Estimation of LD decay in the associated region (0.23 = critical LD). **Step 3:** Extension of the first boundaries using the local LD decay.

The simulation study showed that for small effect loci (5-10 % of the total variance with a trait narrow-sense heritability of 0.6), the proportion of false positive QTL on overall QTL increased by around 40% when the -log(P-value) threshold used to declare a SNP-trait association positive was increased from 3 to 6 (Fig. 9). It can appear counter-intuitive as increasing the -log(P-value) threshold decreases the rate of false

positive SNP-trait associations. In fact, increasing the -log(P-value) threshold decreased false positive QTL from 7.6 to 4.4 % of the total number of tests, but drastically decreased the power of detection (proportion of true positive QTL among the total number of tests) from 71.3 to 28.6 %. Thus, it led to a higher proportion of false positive QTL among all QTL mainly due to a reduction of QTL size (from 7.8 to 4.8 cM) when we increased the -log(P-value) threshold. In continuity, for 32 % of true positive QTL the most significant QTN was not the one closest to the causal mutation. This means that causal mutations are not necessary under significance peak.

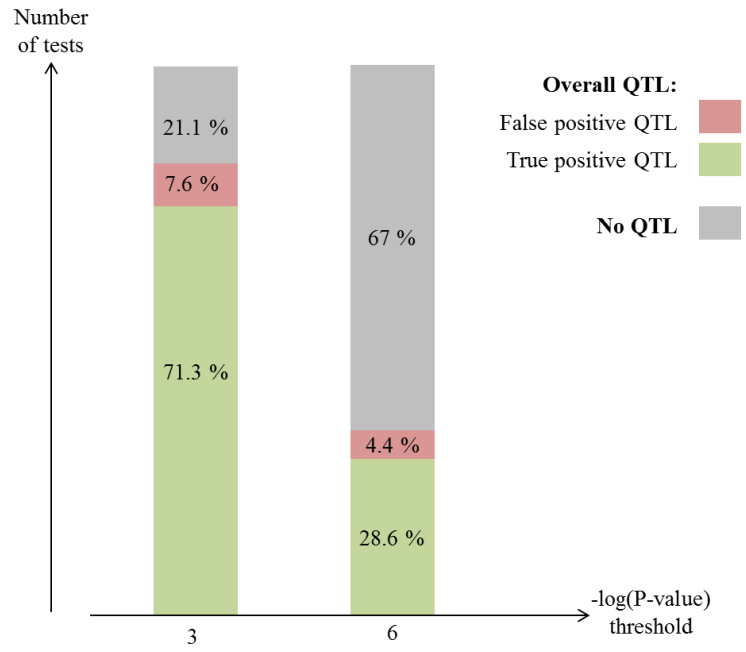


Figure 9: Summary of simulation study results.

Of course, increasing the -log(P-value) threshold decreases SNP-trait false positive rate. Nevertheless, at the end of the QTL definition process, increasing the -log(P-value) threshold does not make our method more efficient.

Gene discovery strategy

In this framework, our gene discovery strategy was not only driven by SNP significance in GWAS. In any case we expected small effect loci and mathematically weak SNP-trait association [small -log(P-value)] as we worked on complex traits. Thus, we choose (i) not to be too stringent on SNP-trait associations even if it may increase mean QTL size and (ii) to prioritize QTL on other criteria (*e.g.* QTL size, location, and previous knowledge on region effects). Indeed, defining QTL boundaries allows for more efficient comparative studies. For example, due to differences of LD structure between panels, LD between genetic markers and causal mutations may vary leading to different QTN between GWAS studies. However, at a larger scale, QTL may be less variable. Synteny approaches will also be more efficient.

As wheat has not been yet completely sequenced, the main issue is to be able to project QTL from a genetic map on a genome sequence. To deal with this issue, Biogemma developed a wheat genome zipper that mimics the wheat genome sequence following Mayer et al. (2011). Thanks to this tool that has repeatedly proven its efficiency; we can readily access to genes under a QTL. With our QTL cloning method, we do not have to create genotypes as we used varieties already available. However, at a given panel, it makes fine mapping of a precise chromosomal region impossible (heterozygotes under QTL are not available). Thus, we may be stuck with "long" QTL containing many genes. Nevertheless, an increase of panel size may decrease QTL length by decreasing LD. Moreover, the quantity of information available (*e.g.* gene annotation, validation in model species, transcriptomic, proteomic and metabolomics datasets) to look for candidate genes is enormous and constantly increasing. Therefore, efforts can be transferred from genotypes creation to data mining.

GWAS combined with a method to define QTL has the potential to speed up QTL cloning process. However, the efficiency of the whole process has to be tested to assess risks and correctly choose the parameters of the method. We decided to apply our strategy to our NUE dataset and published both method and results. The published work is presented in the following part.

References

Choulet F, Alberti A, Theil S, Glover N, Barbe V, et al. (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science DOI: 10.1126/science.1249721

Le Gouis J, Bordes J, Ravel C, Heumez E, Faure S, Praud S, Galic N, Remoué C, Balfourier F, Allard V, Rousset M (2012) Genome-wide association analysis to identify chromosomal regions determining components of earliness in wheat. Theor Appl Genet 124:597-611

Mayer KF, Martis M, Hedley PE, Simkova H, Liu H et al. (2011) Unlocking the barley genome by chromosomal and comparative genomics. Pant Cell 23:1249-1263

Salvi S and Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci. 10:297:304

Tian F, Bradbury PJ, Briwn PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, Mcmullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of maize identifies genes affecting leaf architecture. Nat Genet 43:159–162

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, et al. (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Comm 2:467-477.

PART III: LOOKIN' FOR HOT GENES

Theoretical and Applied Genetics
December 2014, Volume 127, Issue 12, pp 2679-2693
Accepted: 3 October 2014

A GENOME-WIDE IDENTIFICATION OF CHROMOSOMAL REGIONS DETERMINING NITROGEN USE EFFICIENCY COMPONENTS IN WHEAT (*T. AESTIVUM* L.)

Fabien Cormier[1] • Jacques Le Gouis[2] • Pierre Dubreuil[1] • Stéphane Lafarge[1] • Sébastien Praud[1] ✉

[1]Biogemma, Centre de recherche de Chappes, Route d'Ennezat CS90126, 63720 Chappes, France. [2]INRA, UBP UMR 1095, Génétique, Diversité et Ecophysiologie des Céréales, 5 chemin de Beaulieu, F-63 039 Clermont-Ferrand, France.
✉ sebastien.praud@biogemma.com

ABSTRACT: Improving nitrogen use efficiency is a key factor to sustainably ensure global production increase. However, while high-throughput screening methods remain at a developmental stage, genetic progress may be mainly driven by marker-assisted selection. The objective of this study was to identify chromosomal regions associated with nitrogen use efficiency related traits in bread wheat (*Triticum aestivum* L.) using a genome-wide association approach. Two hundred and fourteen European elite varieties were characterised for 28 traits related to nitrogen use efficiency in eight environments in which two different nitrogen fertilisation levels were tested. The genome-wide association study was carried out using 23,603 SNP with a mixed model for taking into account parentage relationships among varieties. We identified 1,010 significantly associated SNP which defined 333 chromosomal regions associated with at least one trait and found colocalisations for 39 % of these chromosomal regions. A method based on linkage disequilibrium to define the associated region was suggested and discussed with reference to false positive rate. Through a network approach, colocalisations were analysed and we highlighted the impact of genomic regions controlling nitrogen status at flowering, precocity, and nitrogen utilisation on global agronomic performance. We were able to explain 40 +/- 10 % of the total genetic variation. Numerous colocalisations with previously published genomic regions were observed with such candidate genes as Ppd-D1, Rht-D1, NADH-Gogat, and GSe. We highlighted selection pressure on yield and nitrogen utilisation discussing allele frequencies in associated regions.

ABBREVIATIONS

ADM_S, straw dry matter at maturity; DArT, diversity array technology; LD, linkage disequilibrium; FLO, flowering date; G, genotype; G × E, genotype × environment; G × N, genotype × nitrogen; GNY, grain nitrogen yield; GPC, grain protein content; GPD, grain protein deviation; GY, grain dry matter yield; HI, harvest index; KS, kernel per spike; N, nitrogen; %N_S, straw nitrogen content at maturity; NHI, nitrogen harvest index; NSA, straw nitrogen per area; NTA, total nitrogen in plant at maturity; NUE, nitrogen use efficiency; NUE_Prot, nitrogen use to protein efficiency; NupE, nitrogen uptake; NutE, nitrogen utilisation efficiency; NutE_Prot; nitrogen utilisation to protein efficiency; P; P-value; PH, plant height; QTL, quantitative trait locus; QTN, quantitative trait nucleotide; SA, spike per area; SNP, small nucleotide polymorphism; SSR, single sequence repeat; TKW, thousand kernel weight

INTRODUCTION

Global production of cereals has increased by around threefold since 1960 (FAO 2012) and is correlated with increased application of nitrogen (N) fertiliser. To date, the global growth in fertiliser demand is still positive as the demand for grain increases (FAO 2011). Thus, to sustainably enhance worldwide cereal production, it is necessary to increase production per N fertiliser unit.

Nitrogen use efficiency (NUE) is defined as grain yield divided by the available nitrogen. In bread wheat (*Triticum aestivum* L.) genetic progress on NUE related traits has been assessed in various studies (Ortiz-Monasterio et al.1997; Guarda et al. 2004; Muurinen et al. 2006; Cormier et al. 2013) and was mainly driven by selection on yield at a constant and high N level. This genetic progress should be at least maintained and preferably accelerated to deal with political, economic, and environmental concerns (Rothstein 2007; Pathak et al. 2011). Several promising ways to improve NUE have been proposed such as focusing on root architecture (Hirel et al. 2007; Foulkes et al. 2009; Kant et al. 2011) or on senescence and remobilisation (Gaju et al. 2011; Distelfed et al. 2014). Although encouraging results have been obtained (Knyazikhin et al. 2013), phenotyping for NUE is still tedious as there are actually no high throughput methods available (Manske et al. 2001; Tester and Langridge 2010). Moreover, G × N interactions have been observed on various agronomic traits (*e.g.* Le Gouis et al. 2000; Barraclough et al. 2010; Cormier et al. 2013) meaning that varieties may have to be tested in several N regimes. Thus, in a global context of fertiliser reduction, the ability to identify stable quantitative trait loci (QTL) controlling NUE related traits and to implement this knowledge in breeding programs may condition a part of the future genetic gain. Various studies have already identified interesting quantitative trait loci (QTL) linked to N metabolism and response to N using biparental populations (*e.g.* An et al. 2006; Laperche et al. 2007; Habash et al. 2007; Guo et al. 2012; Xu et al. 2013). Originally developed in animal and human genetics, genome wide-association study (GWAS) is now used in numerous studies in crop species. Although GWAS has provided useful results in dissecting complex traits in wheat such as yield and its components (*e.g.* Crossa et al. 2007; Neumann et al. 2011), and yield response to nitrogen (Bordes et al. 2013), to our knowledge, this study is the first GWAS on NUE and NUE related traits in small grain cereals.

GWAS overcomes the two main limitations suffered by biparental design of limited allelic diversity and poor mapping resolution due to limited recombination events during the creation of the population (Korte and Farlow, 2013).

Table 1: Description of measured and calculated traits assessed in all environments for which adjusted means by varieties where calculated on a 214 lines wheat association panel.

Trait	Description	Formula	Units	H^2_G	Mean	SD
ABSN	post-anthesis absorption	NTA – NFA	kg ha^{-1}	0.25	22.7	26.43
ADM_FLO	above-ground dry matter at anthesis		kg ha^{-1}	0.69	10618	2222.50
ADM_S	straw dry matter at maturity		kg ha^{-1}	0.84	7288	1861.32
DMGY	dry matter grain yield		kg ha^{-1}	0.89	7400	1257.49
EFFG	genetic efficiency	REMN / GNY	%	0.18	82.3	19.85
EFFREMN	remobilisation efficiency	REMN / NFA		0.27	77.3	7.56
FLO	anthesis date		days (after 1st January)	0.99	149.2	7.12
GNY	grain N yield	GPC / 5.7 × GY	kg ha^{-1}	0.50	127.9	35.44
GPC	grain protein concentration		%	0.92	9.93	2.05
GPD	grain protein deviation	GPC - a × GY - b	% of protein	0.80	0	0.78
HI	harvest index	GY / (GY+ADM_S)	%	0.88	50.4	5.67
INN_FLO	N nutrition index	%N_FLO/(5.35×ADM_FLO/1000)^ (-0.442)		0.63	0.69	0.19
NFA	N at anthesis	ADM_FLO ×%N_FLO	kg ha^{-1}	0.16	138	48.82
NHI	N harvest index	GNY / NTA	%	0.63	81.1	5.71
NSA	straw N per area	ADM_S × %N_S	kg ha^{-1}	0.50	30.4	14.17
NTA	total N per area	NSA + GNY	kg ha^{-1}	0.41	158	45.03
NUE	N use efficiency	GY / NTAmax*	kg DM kg^{-1} N	0.87	37.8	7.69
NUE_Prot	N use efficiency to protein	GPC / NTAmax*	% protein kg^{-1} N ha^{-1}	0.90	0.05	0.01
NupEFlo	N uptake at anthesis	NFA/NFAmax*	%	0.15	0.76	0.12
NupEMat	N uptake efficiency at maturity	NTA / NTAmax*	%	0.37	0.78	0.08
NutE	N utilisation efficiency	GY / NTA	kg DM kg^{-1} N	0.87	48.8	11.19
NutE_Prot	N utilisation efficiency to protein	GPC / NTA	% protein kg^{-1} N ha^{-1}	0.89	0.07	0.01
PH	plant height		cm	0.95	76.6	8.43
REMN	N remobilisation	NFA – NSA	kg ha^{-1}	0.25	109	39.21
SA	spikes per area		nb spike m^{-2}	0.85	412	78.83
TKW	1000-kernel weigth		g	0.96	42.4	4.11
%N_FLO	N concentration at anthesis		%	0.80	1.29	0.34
%N_S	straw N concentration at maturity		%	0.77	0.42	0.13

**NTA_{max} and NFA_{max} are defined as the respective 95^{th} percentile of NTA and NFA (see Cormier et al. 2013)*

However, the use of linkage disequilibrium (LD) to identify marker-trait association at the whole-genome level has also some specific limitations. False positive association (Type I error) can easily arise from population structure. In addition, though the accumulation of recombination allows for a high-resolution mapping, it also decreases LD between causal mutation and markers, which in turn decreases the power of detection for a given number of markers. To deal with these major trade-offs, independent markers can be used to assess the relative kinship in the panel. This information is then used to control Type I error. The power issue can be solved by increasing the number of markers which is now possible with the use of wheat single nucleotide polymorphism (SNP) chips at relatively low cost (Wang et al. 2014).

In GWAS, results are mostly shown using simple Manhattan plots and there is no widespread method to well define associated chromosomal regions.

Indeed, in a panel, the link between linkage disequilibrium and genetic or physical distance is much more complex than in a biparental population, where methods such as one LOD support interval or bootstrapping are commonly used to assess QTL confidence interval (*e.g.* Lander and Botstein 1989; Mangin et al. 1994; Visscher et al. 1996). Moreover, in strong LD regions, pairwise correlation between significant markers can approach genotyping accuracy rate. Thus, even with methods such as stepwise logistic regression to test whether a marker in a given set is necessary or sufficient to explain the association signals, finding the one likely to be closest to the causal mutation is nearly impossible (McCarthy and Hirschhorn 2008). Added to that, in high LD regions, the tested marker is correlated to many other SNPs that can contribute to the estimation of the kinship reducing the power of detection (Rincent et al. 2014). Thus, the most significant quantitative trait nucleotide (QTN) may not be the closest to the causal mutation. In low LD regions, it is possible that only one SNP is significant, and there is no simple way to define a region from the relationship of P-value (P) with genetic/physical distance. In any case, P-value depends on the QTL effect. This biases the P-value support method of constructing "confidence interval" (Mangin et al. 1994). Thus, authors often fix a more or less arbitrary window around QTN peaks based on mean LD decay, for example 1 Mb in maize for Tian et al. (2011), 200 kb in rice for Zhao et al. (2011), or 5 cM in wheat for Le Gouis et al. (2012). The method chosen to define an associated chromosomal region influences GWAS reliability and this issue remains under investigated.

Using 214 European elite varieties, 28 NUE-related traits, and 23,603 SNP, this study aimed to (i) estimate the power of such an elite panel to perform GWAS with respect to the method used to define associated chromosomal regions and false positive rate, (ii) identify stable chromosomal regions involved in NUE related-traits and assess their transferability to the field, and (iii) analyse colocalisations for NUE components and NUE related traits to estimate pleiotropic effects associated with QTL-based selection.

MATERIALS AND METHODS

Phenotypic data

Phenotypic data are described in Cormier et al. (2013). Briefly 225 European elite varieties were evaluated in eight environments defined as a combination of year, site, and nitrogen supply (two seasons, three sites, and two nitrogen supplies). The high N treatment corresponded to common agricultural practices. The low N treatment corresponded to a mean yield reduction of 20% (Suppl. data 1). Other crop inputs including weed, disease and pest control, potassium, phosphate and sulphur fertilisers, were applied at sufficient levels to prevent them from limiting yield. Plant growth regulators were applied to limit lodging in all environments. In each environment, 28 traits were measured or calculated (Table 1). From adjusted means by trial, overall adjusted means by varieties were computed using a simple linear model with environment and genotype as fixed effects. These values were used in the GWAS. Generalized broad-sense heritabilities (H_G^2) were calculated using the formula proposed by Cullis et al. (2006) from the

previous linear model with genotype as a random effect.

Genotyping and consensus map

Of the 225 varieties present in field trials, 214 were genotyped. SNP data consisted of a subset of SNP from an Illumina 90K chip (Wang et al. 2014) together with SNP developed by Biogemma. Heterozygous loci were considered as missing data. Loci with a minor allele frequency inferior to 0.05 or loci which had available data for less than 150 varieties were not used. In total, we used 23,603 mapped SNP in this study.

We built a consensus map with the Biomercator software (Arcade et al. 2004). We used the map published by Le Gouis et al. (2012), based on Somers et al. (2004), as a reference. This map contains SSR and DArT markers, and the location of several major genes (Vrn, Ppd, Rht). SNP were projected on it, from non-published maps containing 535 markers in common with this reference map. The Strudel software was used to check map alignments (Bayer et al. 2011) and mapping errors were corrected.

Linkage disequilibrium

We used the r^2 estimator (Hill and Robertson, 1968) to assess linkage disequilibrium (LD). LD was calculated for every pair of markers mapped on the same chromosome, and then r^2 was plotted against map distance. For every chromosome, LD decay (cM) is estimated at the point where a curvilinear function proposed by Hill and Weir (1988) intersects the threshold of the critical LD. Critical LD was the 95th percentile of the unlinked-r^2 assessed on 100,000 randomly chosen pairs of unlinked loci (mapped on different chromosomes) which were square root transformed to approximate a normally distributed random variable (Breseghello and Sorrells 2006).

Association mapping study

Following Patterson et al. (2006), we did not find any structure in this 214-varieties panel. Indeed, the largest eigenvalue was not significant (P=0.043). Thus, we tested SNP-trait association using a mixed model K (Yu et al. 2006) written in R using the ASReml-R package (Butler et al. 2007) and expressed as:

$$y = 1\mu + \mathbf{S}\alpha + \mathbf{Z}u + \varepsilon$$

where y is a vector of estimated genetic values, 1 is a vector of 1's, μ is the intercept, α is the additive effect of the tested SNP, u is a vector of random polygenic effects assumed to be normally distributed $N(0, \sigma_y^2\mathbf{K})$ with **K** a matrix of relative kinship, **S** and **Z** are incidence matrices, ε is a vector of residual effects.

K was estimated as 1(n × n) - Rdist where Rdist is the modified Rogers' distance (Rogers 1972) matrix based on 3 461 SNP spread over the genome and with less than 0.1 missing data and 1(n × n) is a matrix of 1's of the same size as the Rdist matrix (n = 214).

To summarise, we tested 23,603 SNP on 28 traits using the adjusted means of 214 European elite varieties. There is no widespread method to define QTL boundaries from GWAS results. So, we proceeded as follows. First, for each trait, we computed LD between every significantly associated SNP (quantitative trait nucleotide - QTN). LD blocks were defined as a group of QTN

belonging to the same LD cluster (clustering by average distance) using a cutoff of (1-"critical LD"). We define the initial QTL boundaries as the minimum and maximum map position of QTN belonging to the same LD block. Then, as previously described, we assessed LD between every mapped SNP within a window covering 10% of the chromosome length and centred on each QTL. We used the LD decay to extend the previous boundaries. This second step aimed to take into account possible LD with the causal mutation at the first QTL boundaries (for detail Suppl. data 2). We only defined QTL for LD blocks containing SNP mapped on the same chromosome. For each trait, QTL with overlapping boundaries were considered the same if the alleles increasing the trait value at each were themselves correlated positively.

Phenotype simulation and power

The statistical power provided by the panel was evaluated through simulation studies where -$\log_{10}$(P) thresholds, narrow-sense heritability and variance explained by a SNP were the three modulated parameters. We set -$\log_{10}$(P) threshold at 3, 4, 5, 6; narrow-sense heritability (h^2) at 0.3, 0.6, and 0.9; and variance explained by the SNP (π) at 0.010, 0.030, 0.050, 0.075, 0.100, 0.150, and 0.200. Phenotypes were simulated as follows:

$$y_i = g_i + a_{ij} + \varepsilon_i \qquad (1)$$

where y_i is the simulated phenotype of the variety i, g_i is the genetic additive background effect of variety i, a_{ij} the additive effect at the quantitative trait nucleotide (QTN) j of variety i allele, and ε_i a residual error term sampled from a normal distribution $N(0, \sigma\varepsilon^2)$.

First, k=100 SNP were chosen to simulate the genetic background effect. This selection is made by forming *k*-means cluster based on the genotyping incidence matrix and selecting the SNP nearest the centroid of each cluster (Lorenz et al. 2010). Thus, if g_i is the genetic background effect of variety i:

$$g_i = \sum_{k=1}^{k=100} a'_{ik} \, , \; a'_{ik} = \begin{cases} 1 \\ 0 \end{cases} \qquad (2)$$

with a'_{ik} the effect of the variety i allele at the locus k.

Narrow-sense heritability (h^2) is defined by:

$$h^2 = \frac{\sigma g^2 + \sigma j^2}{\sigma T^2} \qquad (3)$$

where σj^2 the genetic variance related to QTN j different from k, σg^2 the variance related to the genetic background, and σT^2 the total variance.

The variance explained by QTN j (π) is defined by:

$$\pi = \frac{\sigma j^2}{\sigma T^2} \qquad (4)$$

Total variance (σT^2) is deduced from equation (3) and equation (4) as h^2 and π are fixed in each simulation study:

$$\sigma T^2 = \frac{\sigma g^2}{h^2 - \pi} \qquad (5)$$

Given the percentage of variance explained by QTN j (π), its additive effect (aj) is calculated by Falconer and Mackay (1996) as:

$$a_j = \sqrt{\frac{\pi \times \sigma T^2}{p_j(1-p_j)}} \qquad (6)$$

with p_j the allele frequency of the reference allele at locus j. Thus, if variety i allele at QTN j was the reference allele, a_{ij} from equation (1) was equal to a_j, else a_{ij} was equal to $-a_j$.

Finally, the variance of the residual error term ($\sigma\varepsilon^2$) was computed as:

$$\sigma\varepsilon^2 = (1 - h^2) \times \sigma T^2 \qquad (7)$$

In total 400 SNP were randomly chosen to play in turn the role of the QTN j with $j \neq k$ (QTN ≠ genetic background effect) for each pair of h^2 and π parameter values. The statistical model used to detect associations between SNP and simulated phenotypes was the previously described model K. In the same way, QTL were defined following the two steps already described. Detection power was estimated by the ratio of the number of times a true QTN was located in the computed QTL to the total number of tests. The SNP selected as being the true QTN j was not tested *per se*.

Prediction

The percentage of total variance explained by each significant SNP was first assessed for each trait using a simple regression of overall adjusted mean on the SNP (r^2_{snp}). Then, for each trait, the predicted values of varieties were estimated by summing the allele effects assessed in GWAS at associated loci. To avoid redundancy, only one SNP per LD block was kept; that which explained the most variance.

This model was first used to predict overall adjusted means. It was then used to predict adjusted means in each of the eight individual environments. Consequentially, we computed two types of correlations (r^2): the correlation between predicted values and overall adjusted means (r^2_{adj}), and the correlation between predicted values and each of the eight individual environments (r^2_{env}).

To assess transferability of GWAS results to field trials, we calculated a prediction similarity [mean(r^2_{env})/r^2_{adj}] that we plotted as a function of trait heritability.

Colocalisation and network approach

To assess the impact of genetic correlation and pleiotropy, we analysed colocalisations through a network approach. QTL colocalisation between two traits were statistically tested using the probability of an hypergeometric law ("sampling without replacement"; Larsen and Marx, 1985) with the total cumulative length of QTL for trait i and trait j and the total map length as parameters of the hypergeometric distribution. The cumulative length of QTL shared by trait i and j was the parameter of the probability. A fairly stringent threshold of P = 0.001 was set as the criteria of significance.

On the basis of significant colocalisations, inter-trait relationships were then studied through a network approach using traits as nodes and the percentage of one trait QTL overlapping another trait QTL as edges. Betweenness centrality was computed on each node following Opsahl et al. (2010) method with $\alpha = 0.5$ to equally take into account the number of edges and edges' weights in the calculation. To statistically test trait betweenness centralities values, this network was then permuted 500 times to assess the empirical distribution of betweenness centrality, and thus

determine the statistical law underlying this distribution.

RESULTS

Genetic map and linkage disequilibrium

Table 2: SNP used in association: number of mapped SNP, coverage on the consensus map, SNP density and LD decay at a critical LD r^2 = 0.23. Critical LD was assessed as in Breseghello and Sorrells (2006).

Chr	SNP	Coverage (cM)	SNP density (cM^{-1})	LD decay (cM)
1A	1246	110.4	11.3	0.49
1B	2,055	128.5	16	0.19
1D	430	121.7	3.5	2.71
2A	1,454	262.7	5.5	1.39
2B	,2362	205.8	11.5	0.70
2D	402	130.9	3.1	0.80
3A	1,151	155.1	7.4	0.68
3B	1,972	147.8	13.3	0.05
3D	253	104.7	2.4	1.07
4A	786	123.4	6.4	0.21
4B	849	143.3	5.9	0.70
4D	97	139.7	0.7	2.43
5A	1,604	186.1	8.6	0.32
5B	2,243	262.4	8.5	2.19
5D	327	115.6	2.8	0.94
6A	1,588	122.0	13	0.19
6B	1,603	115.0	13.9	0.05
6D	254	136.8	1.9	1.02
7A	1,782	122.2	14.6	0.38
7B	1,034	198.5	5.2	1.06
7D	246	134.9	1.8	6.00
Total	23,603	3,167.5	7.5	1.12

The consensus genetic map obtained had a total length of 3,167 cM. To finely map QTL, LD has to decay rapidly and SNP density has to be high to ensure that at least one SNP is linked to the causal mutation. While diversity level is similar in the A and B genomes, it is greatly reduced in the D genome (Cadalen et al. 1997), contributing to its higher levels of LD.

Indeed, mean LD decay on genome A, B, and D was respectively 0.52, 0.70, and 2.14 cM. LD decay is the estimated distance from which two SNP are not genetically linked, meaning that their LD (r^2) is inferior to the critical LD. Critical LD was estimated from a sample of 100,000 pairs of unlinked SNP which revealed a mean unlinked-r^2 of 0.016 and a critical LD (95^{th} percentile) of 0.23.

A rapid LD decay predicts a good mapping resolution in GWAS. Though as previously mentioned, it can decrease power if SNP density is not sufficient. SNP density ranged from 0.7 cM^{-1} for chromosome 4D to 14.6 cM^{-1} for chromosome 7A (Table 2). On genomes A and B, SNP density seemed sufficient with respect to LD decay. On genome D, the lower SNP density may be compensated for by the higher LD, but QTL will be less precisely defined.

Power assessment

Choosing a P-value threshold has to balance the control of Type I error (false positive) with Type II error (false negative). Considering power simulation and the expectation of small effect QTN, a $-\log_{10}(P)$ threshold of 3 was adopted as a criterion for significant marker-trait associations. Indeed, a more stringent threshold inflated Type II error and thus reduced extremely the power of detection, notably on QTN explaining less than 10% of the variance (Fig. 1).

At a QTN heritability of 5 % and a narrow-sense heritability of 0.6, power was dramatically reduced from 55 % to 7 % when $-\log_{10}(P)$ threshold increased from 3 to 6 (Fig. 1).

Table 3: QTL detected on a wheat association panel for 28 traits. QTL boundaries were defined as the minimum and maximum genetic position of QTN belonging to the same LD block (for LD blocks containing SNP mapped on the same chromosome) extended by the LD decay assessed on a window covering 10 % of the chromosome length centered on the mean genetic position. See Table 1 for trait abbreviations.

Trait	QTL		Effect (%)[b]		MAF[c]		Size (cM)		Group 1			Group 2			Group 3			Group 4			Group 5		Group 6			Group 7		
	Total	Positif QTL[a] (%)	mean	SD	mean	SD	mean	SD	A	B	D	A	B	D	A	B	D	A	B	D	A	B	A	B	D	A	B	D
ABSN	13	69	12.9	2.2	0.20	0.12	5.95	11.50		2		1	1			1		1	1		1	1	1	1			2	
ADM_FLO	12	50	2.1	0.5	0.23	0.12	2.14	2.71	1		1		2	1	1	1		1						1		2	1	
ADM_S	16	50	2.7	0.5	0.26	0.13	4.12	7.67		2			1	2	1	1			1	1	2	1	1			1	1	1
DMGY	10	70	2.3	0.7	0.19	0.13	0.79	0.82		1			2		2						1		2				1	1
EFFG	19	32	2.4	0.5	0.24	0.11	3.27	8.27	2	1		2	3			1		2			1	1	1	1		1	2	1
EFFREMN	12	67	1.2	0.3	0.18	0.12	2.33	3.00	1	2			1				2	1			1		1				3	
FLO	18	78	0.8	0.2	0.28	0.13	1.64	1.66	1			1	4	3	1	1					4		1		1	1		
GNY	11	36	1.5	0.3	0.28	0.12	6.96	9.07		3	1	1			1				1		1		1			1	1	
GPC	8	13	3.0	1.2	0.19	0.14	4.34	8.43					1		1	1		1			1	1			1		1	
GPD	8	38	0.16	0.04	0.31	0.15	1.63	1.16				2	2										2		1	1		
HI	18	72	1.6	0.6	0.25	0.16	1.73	2.20	1	1		2	2	1	1	3		1	1		1	1	1			1	1	
INN_FLO	7	14	2.1	0.2	0.27	0.09	4.86	6.40		1			2	1								2	1					
NFA	10	10	2.1	0.6	0.21	0.08	1.93	3.56					2		1	1		2			1	1		1				1
NHI	10	80	0.6	0.1	0.29	0.14	2.52	3.97	1	2		1							1		2	1				2		
NSA	14	43	3.4	1.4	0.24	0.16	1.62	2.81		1		1	3	1		2	1	1			1	1				1	1	
NTA	8	13	1.5	0.3	0.25	0.16	5.66	8.56		1	1	1	1								1		2			1		
NUE	14	57	2.3	0.6	0.22	0.12	3.00	6.61	1				1	1	2	2	1	1			1				1	2		1
NUE_Prot	11	18	2.9	1.1	0.22	0.15	4.30	7.13					2	1	1	1		1			1	2			1		1	
NupEFlo	7	0	2.2	0.4	0.22	0.11	2.36	4.23				1	1		1			1				1		1		1		
NupEMat	10	30	1.5	0.3	0.27	0.15	5.38	7.80		2	1	1						1			2	1	1			1		
NutE	6	67	2.6	0.6	0.22	0.13	1.22	1.73					1	1	1			1				1			1			
NutE_Prot	16	25	2.5	0.9	0.23	0.13	2.47	6.20	1				2		2	1	1	1			3		1		1	1	1	1
PH	14	14	2.9	0.5	0.20	0.11	7.67	14.97		1			2			1		1		1	2	1	2			2	1	
REMN	12	42	2.7	0.4	0.20	0.11	1.80	4.06		3			1					1				1	1			2	2	1
SA	11	45	4.3	1.2	0.15	0.12	1.44	1.61					1		2	2		1					2	1	1		1	
TKW	9	44	2.4	1.0	0.25	0.13	3.40	3.37		2				1		1		1				1	2				1	
%N_FLO	8	25	2.6	0.5	0.31	0.13	2.24	2.40					2	1								1	3	1				
%N_S	21	33	4.0	1.0	0.19	0.10	2.97	4.58		3		3	1	1	1	2		1	1	1		3	2				2	
Total									9	28	4	17	41	15	19	22	5	21	6	3	27	22	28	7	8	21	23	7

[a] *Percentage of QTL for which the most frequent allele had a positive effect on trait*
[b] *Effect expressed in percentage of trait mean (except for GPD)*
[c] *MAF = minor allele frequency*

At a -$\log_{10}$(P) score threshold of 3, when the genetic variance explained by the locus was greater than 10 %, trait heritability did not affect power and Type II error was reduced. In general, the variance explained by the QTN was the main factor that influences the power of the study as compared to trait narrow-sense heritability. It should be noted that with a weakly stringent threshold of 3 the power to detect an association for a QTN, which explained 5 % of the total genetic variance was 48, 55, and 60 %, for a trait narrow-sense heritability of 0.3, 0.6, and 0.9, respectively.

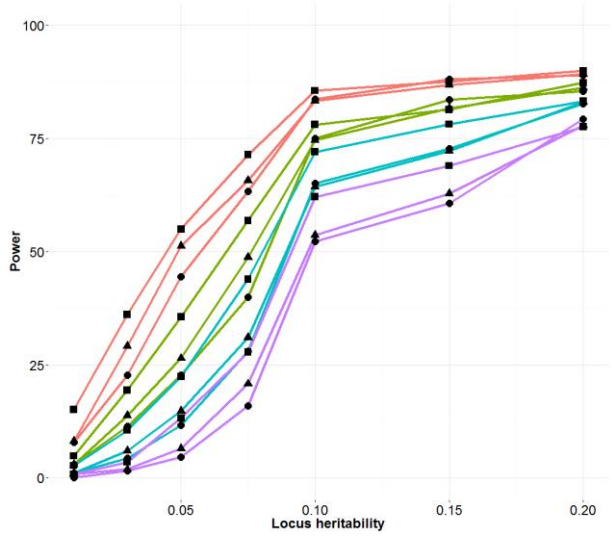


Figure 1: Influence of trait heritability and -log10(P-value) threshold on the relation between locus heritability and power of detection in a 214-lines wheat association panel. In red, green, blue, violet, respective LOD score thresholds are 3, 4, 5, and 6. Square, triangle, and circle represent a respective narrow-sense heritability of 0.9, 0.6, 0.3.

GWAS results

Overall, 1,010 SNP were significantly associated (QTN) to at least one of the 28 studied traits. Considering QTN LD blocks and LD around associated regions, 333 QTL were mapped with a mean size of 3.2 cM. Ninety percent (between the 5th and 95th percentile) of QTL had a range within 0.1-14 cM indicating that the method used to define QTL is mostly efficient. In few cases, the assessments of LD decay in the chromosomal region containing QTN may not correctly fit and QTL boundaries must be used with caution.

In agreement with SNP density and the genetic diversity, the number of QTL on genome D (42) was smaller than on genome A (142) and B (149). Homeologous group 2 maximised the number of QTL with 73 QTL. The number of QTL by trait ranged from 6 for NutE to 21 for %N_S (Table 3).

Predictions

First, we assessed the variance explained by each significant SNP (QTN). Then, we predicted overall adjusted means and each of the eight environments' adjusted means. On average, QTN explained 8.81 +/- 4.79 % of the overall adjusted means (r^2_{snp}). On overall adjusted means, the best prediction (r^2_{adj}) was made on HI (Table 4). Using 20 SNP, we were able to explain 61.4 % of the genetic variation. Using 15 SNP on NUE, we were able to explain 55.7 % of the overall adjusted mean variation (Fig. 2) and 29.7 +/- 4.9 % of the individual environment's variation (Table 4). On the environments' data (r^2_{env}), flowering date was the best predicted trait with 55.3 % of the variation explained on average.

Differences between predictions made on overall adjusted means (r^2_{adj}) and predictions on individual environment values (r^2_{env}) resulted from genotype × environment interactions. Thus, it was linked to trait broad-sense heritability. In fact, the transferability of our GWAS results to environmental values was exponentially proportional to trait broad-sense heritability (Fig.

3). This means that GWAS results became rapidly powerless to predict phenotypic values as broad-sense heritability decreased.

Table 4: Summary of GWAS results predictions made by SNP (r^2snp) and using the sum of SNP effect on both overall adjusted mean (r^2adj) and on eight individual environments (r^2env). To avoid redundancy, for each LD block, the SNP which maximized the genetic variance explained was selected.

Trait	SNP[a]	Prediction on adjusted means			Prediction on individual environments	
		r^2_{snp} (%)		r^2_{adj} (%)	r^2_{env} (%)	
		mean	sd		mean	sd
ABSN	14	6.0	0.9	37.7	6.9	3.3
ADM_FLO	13	6.9	4.4	40.9	18.5	13.1
ADM_S	17	6.5	3.8	52.8	27.7	4.2
DMGY	12	11.5	9.0	53.6	30.8	6.2
EFFG	20	6.1	1.0	42.3	7.0	3.6
EFFREMN	13	7.5	1.7	40.4	8.4	4.4
FLO	20	8.6	6.5	58.5	55.3	2.6
GNY	11	7.3	2.9	40.0	9.9	5.7
GPC	10	14.0	8.7	57.5	37.9	10.8
GPD	8	7.8	3.9	33.7	15.6	5.1
HI	20	8.6	6.4	61.4	32.4	4.3
INN_FLO	8	11.5	4.3	40.0	12.8	10.5
NFA	13	6.3	2.3	34.2	5.7	5.2
NHI	11	5.3	2.9	37.2	10.9	5.8
NSA	15	6.3	3.4	38.2	9.7	5.2
NTA	9	8.1	3.0	32.0	7.0	6.1
NUE	15	8.7	7.2	55.7	29.7	4.9
NUE_Prot	11	12.4	8.8	59.7	35.5	11.5
NupEFlo	9	7.4	3.0	27.7	5.2	5.6
NupEMat	11	6.4	2.9	31.4	6.9	4.3
NutE	6	8.7	6.4	38.3	23.2	9.1
NutE_Prot	18	10.1	8.7	59.8	34.4	7.4
PH	17	10.5	4.9	48.6	37.0	16.0
REMN	12	6.3	1.4	28.3	4.8	3.5
SA	12	7.4	3.8	41.0	22.1	8.1
TKW	10	8.1	2.9	39.0	32.3	3.6
%N_FLO	10	11.4	6.9	45.5	20.3	8.6
%N_S	21	8.3	4.4	57.8	25.8	13.5

[a] SNP number can differ from QTL number in Table 3 when LD blocks contained SNP mapped on different chromosomes (as no QTL was defined but one SNP was used in prediction).

Colocalisation network

Altogether, the QTL covered 20 % (646 / 3,167) of the genetic map. There were colocalisations for 39 % of the QTL identified. Major regions of colocalisation were on chromosomes 1B, 2B, and 7A (Suppl. data 3). Considering NUE and its two components, N uptake and N utilisation, there was no common QTL between NupEMat and NUE, but two NutE QTL (out of six) colocalised with NUE QTL and acted in the same way on both traits. NUE QTL (9/14) which colocalised with NutE_Prot QTL had opposite effect on these traits. By comparing QTL for the N uptake efficiency at flowering time (NupEFlo) and at maturity (NupEMat), we found that only one QTL was in common between these two traits.

Figure 4 provides a visual representation of the frequencies of QTL colocalisations. Using a bootstrap procedure with 500 permutations, it was assessed that the empiric betweenness centrality followed a gamma distribution (shape = 2.169, rate = 0.079; Suppl. data 6). This distribution was used to test trait betweenness centrality. Four traits had a significant ($P < 0.05$) high betweenness centrality: INN_FLO, FLO, NutE, %N_Flo were ordered from the most significant to the less significant. We should notice that INN_FLO, %N_S, and FLO were not independent as we detected four chromosomal regions of colocalisations between these three traits. Two of them affected the three traits in the same ways. Two of them acted oppositely between FLO and the two other traits. All common QTL between %N_Flo and INN_FLO affected both traits in the same way.

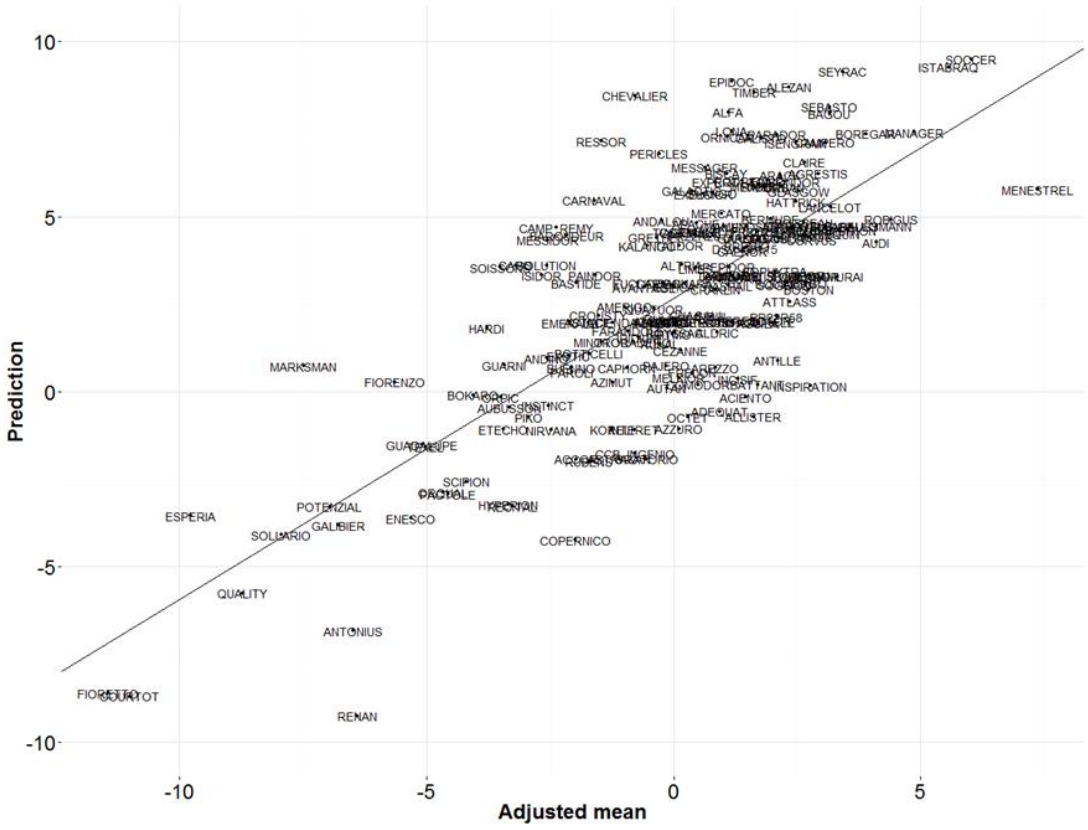


Figure 2: Prediction of NUE values as a function of overall adjusted mean for 214 wheat lines. Predictions were made summing the effects of 15 significantly associated SNP. The following regression function is also plotted: $y = 0.86x + 2.66$ ($r^2 = 0.56$; $P < 0.001$).

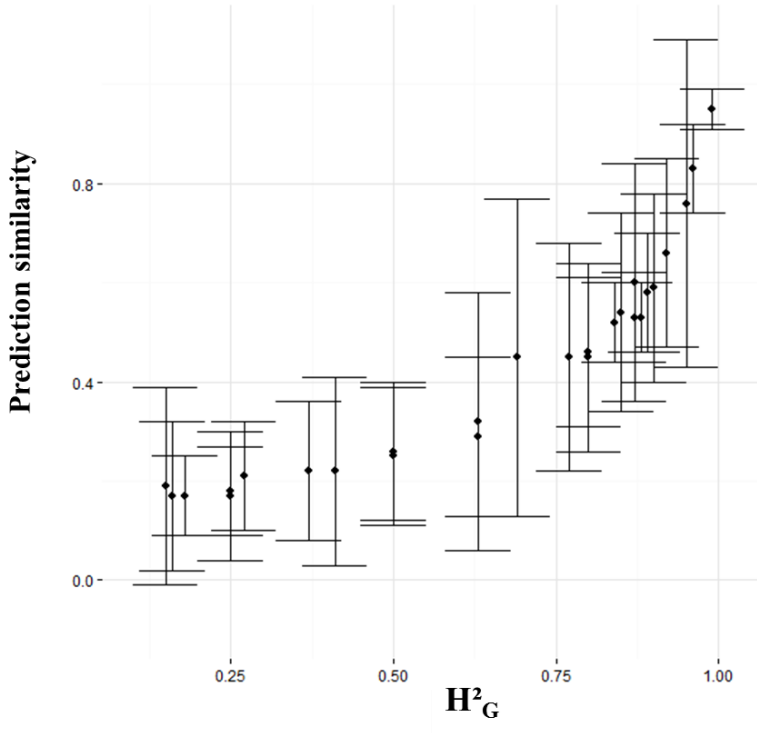


Figure 3: Prediction similarity (r^2env / r^2adj) between predictions made on overall adjusted means (r^2adj) and the ones made on individual environments values (r^2env) as a function of generalized heritability (H^2G) of 28 traits. Means (diamond), standard deviations (whisker). Mean (r^2env/r^2adj) = -0.39eH^2G ($r^2 = 0.88$; $P < 0.001$).

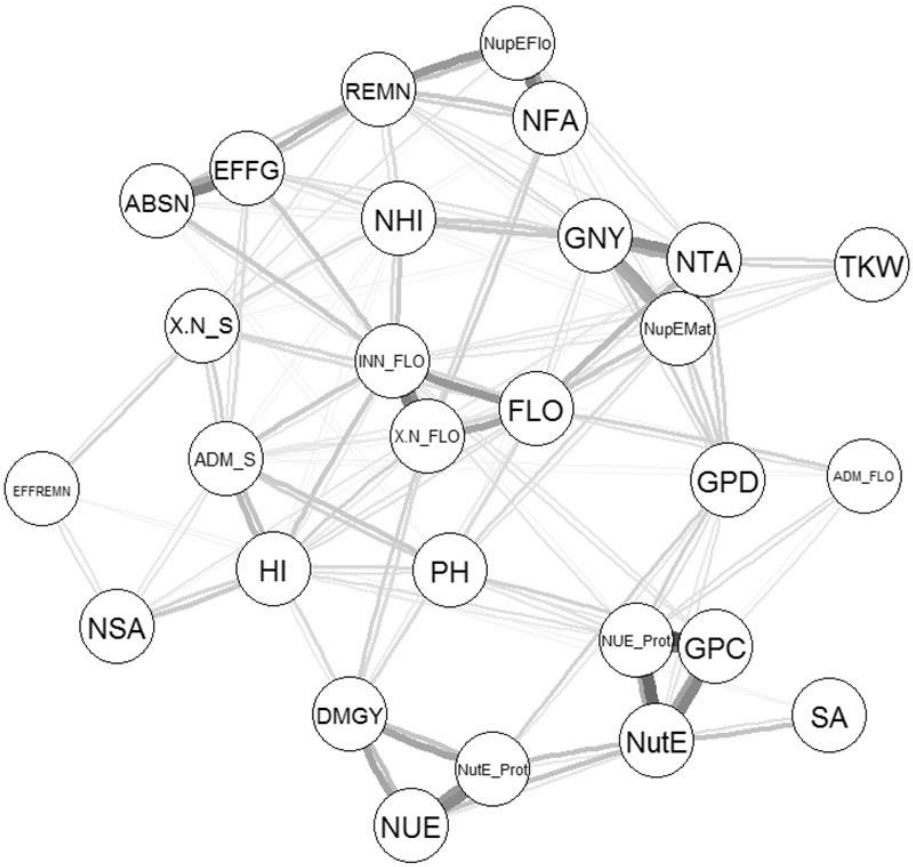


Figure 4: Network of QTL colocalisations for 28 traits measured on a 214 line wheat association panel. This network is based on the percentage of common QTL between traits after correction using a hypergeometric law to determine significant colocalisations ($P < 0.001$). Link thickness is function of the percentage of common QTL, from 5 % for the thinnest to 100 % for the thickest (values in Suppl. data 5).

DISCUSSION

QTL definition and power

In most studies, authors fixed a window around QTN peaks often based on linkage disequilibrium to define associated chromosomal regions in GWAS. However massive variation of LD exists along the chromosomes in wheat (Würschum et al. 2013). In this study, we suggested a method based on LD between QTN and LD within the chromosomal region of interest and assessed its power of detection. This method had the advantage of being based on LD decay in the chromosomal region of interest. Moreover, authors focus on P-value methods (*ad hoc* and *post hoc*) to control false positive rate, although the way they design their associated region influences it. Indeed, linkage disequilibrium between causal mutations and associated SNP or mapping error can lead to the construction of a chromosomal region which does not contain the causal mutation even though the SNP-trait association was real.

Regarding power simulation and error type II, we chose a $-\log_{10}(P)$ threshold of 3 to validate SNP-trait associations. Our real false positive rate (error

type I) was not only influenced by this $-\log_{10}(P)$ threshold. Indeed, in our real error Type I, we should consider all QTL which did not contain the causal mutation whether the SNP-trait association was real or not. Using the results of the power simulation studies we estimated our real false positive rate at 7 % (for a QTN heritability between 5 and 10 %; Suppl. data 2). If we had chosen a $-\log_{10}(P)$ threshold of 6, it would have been 3 %. Thus, increasing P-value threshold reduced real error Type I for small effect QTN yet drastically decreased power (Fig. 1). Moreover, for QTN with a heritability > 10 %, a P-value threshold superior to 3 slightly increased the real error Type I due to smaller QTL (Suppl. data 2).

In GWAS, the real issue to control error Type I is not in the definition of a stringent P-value threshold. It is in the development of a powerful method to define QTL boundaries, particularly in the case of GWAS oriented to gene discovery. This field has practically never been investigated and publications mainly focus on P-value. We advocate balancing QTL coverage, real error Type I, and power altogether. An improvement of our methods could be to adapt the construction of the associated region to QTN heritability.

Power, locus heritability, and genetic determinism

The fraction of total genetic variance explained by a single significantly associated SNP (QTN) averaged 8.81 +/- 4.79 %, which is coherent regarding the simulation study. Indeed, the power started to be maximised from a locus heritability of 10 % (at a $-\log_{10}(P)$ threshold = 3, Fig. 1). Yet variability existed and fraction of total genetic variance ranged from GPC (14.0 +/- 8.7 %) to NHI (5.3 +/- 2.9 %). When numerous QTN explained a small fraction of genetic variance, we can presume that the GWAS study was powerful and that the genetic determinism underlying this trait is highly polygenic. When QTN have larger locus heritability, the cause can be a less polygenic genetic determinism and/or a lack of power due to low narrow-sense heritability. Narrow-sense heritability estimates the proportion of additive variance on total variance (Falconer and Mackay 1996). Thus, narrow-sense heritability is also linked to the importance of epistasis in the trait genetic architecture. In this study we have not searched for epistasis. However, several studies have highlighted its impact. For example, GPC is controlled by major protein concentration genes (Payne 1987; Uauy et al. 2006; Avni et al. 2013) and significant interactions between them (Dumur et al. 2004; Conti et al. 2011; Plessis et al. 2013). Another example is epistatic contribution in the genetic control of PH is important and revealed by Novoselovic et al. (2004), Zhang et al. (2008), and Wu et al. (2010). Using a doubled haploid wheat population, Zhang et al. (2008) estimated firstorder epistatic contribution up to 19.9 % of the PH phenotypic variation.

Authors have often focused on epistatic interactions between SNP having a significant additive effect. However epistatic interactions between SNP without additive effect can also explain genetic variability (Huang et al. 2014) as detected for heading date (Le Gouis et al. 2012). Nonetheless whole genome scan for epistasis is a real computational and analytic challenge, which will

surely help pathways mining (Philipps 2008; Mackay 2014).

Candidate genes and comparison with previously published QTL

Altogether, we detected 333 QTL on 28 traits. Significant colocalisations (QTL boundaries overlapping) between some of them and candidate genes or previously published QTL deserve to be pointed out. Regarding major genes for precocity, only the photoperiod sensitivity gene Ppd-D1 on chromosome 2D colocalised with QTL of FLO, HI, INN_FLO, %N_FLO, %N_S, affecting all these traits in the same way (late genotype have higher HI, INN_FLO, %N_FLO, and %N_S). Ppd-D1 also colocalised with an ADM_S QTL, with an opposite effect. Two factors can explain that Vrn genes were not associated to precocity: (i) this panel contains only winter wheat varieties and (ii) only autumn trials were sown with vernalization requirements fulfilled.

On chromosome 4D, the dwarfing gene Rht-D1 (Rht2) was tested and had an expected significant effect on PH and ADM_S.

Similarly, the three closely mapped genes coding the glutenins and gliadins (Glu3A, Glu3B, and Gli) not surprisingly colocalised with QTL of NUE and NutE_Prot located on chromosome 1A. Moreover, the structural gene for high molecular weight glutenins GluD1 located on chromosome 1D lay within the boundaries of QTL affecting GNY, NTA, and NupEMat.

Several genes from the N assimilation pathway have already been associated to NUE QTL including the genes coding for glutamate synthase (NADH-Gogat) located in QTL on chromosome 3A, and 3B (Quraishi et al. 2011). On chromosome 3A, this colocalised with QTL of NFA, NupEFlo, and %N_S. On chromosome 3B, the NADH-Gogat gene colocalised with QTL of NUE_Prot, GPC, and ABSN. The gene for glutamine synthetase GS1 on 6A (Habash et al. 2007) colocalised with a cluster of QTL for EFFREMN, GPD, NutE_Prot, DMGY, and %N_S. Several publications already mentioned this region as affecting grain number per ear (Habash et al. 2007; Quarrie et al. 2005), NupEMat (An et al. 2006; Xu et al. 2013), root dry weight (An et al. 2006), %N_S and DMGY (Xu et al. 2013).

On chromosome 4B, a QTL of %N_S colocalised with numerous previously published QTL of nitrogen efficiency related trait (An et al. 2006; Guo et al. 2012), glutamate dehydrogenase and glutamine synthase activity (Fontaine et al. 2009), harvest index (Xu et al. 2013), ears, spike, and grain related trait (Quarrie et al. 2005; Habash et al. 2007; Laperche et al. 2007; Fontaine et al. 2009), and root morphology (Laperche et al. 2006). Previously published results were in part due to the presence of Rht-B1 (Rht1) in this chromosomal region. In our case, a diagnostic marker for Rht-B1 was tested and no significant effect was detected for any trait most probably because of the unbalanced allele frequencies of the combination of Rht-B1 and Rht-D1 (0.05, 0.65, 0.18, and 0.12 for the four allelic classes *Rht-B1b/Rht-D1b*, *Rht-B1b/Rht-D1a*, *Rht-B1a/Rht-D1b*, and *Rht-B1a/Rht-D1a*). The glutamine synthetase gene GSe (Habash et al. 2007) mapped using the SSR gpw7026 (Sourdille et al. 2004; Fontaine et al. 2009) was also within this QTL confidence interval and may be a good candidate gene to investigate.

On chromosome 2A, the Rbcs (Xpsr109) gene for the small subunit of the chloroplast photosynthetic enzyme ribulose-1,5-bisphosphate carboxylase / oxygenase (Rubisco) was located in a %N_S QTL, and has already been shown to colocalise with a QTL for N grain concentration (Laperche et al. 2006), and from a meta-QTL analysis on yield and yield-related traits (Zhang et al. 2010). Considering the small size of this QTL in this study (1.6 cM), and the link between N remobilisation and Rubisco subunit expression and degradation (Hörtensteiner and Feller 2002; Gregersen et al. 2008), Rbcs has to be considered as a good candidate gene.

Further investigations are needed on two promising regions where no obvious candidate genes were found within QTL boundaries. On chromosome 5B (gwm67-BCD351), a region linked to the INN_FLO colocalised with QTL previously published by Fontaine et al. (2009) on carbon percentage in flag leaf, and Habash et al. (2007) on nitrogen percentage in peduncle. As the nitrogen nutrition index (INN) refers to the minimum N concentration enabling maximum biomass growth (Justes et al. 1994) this confirms the effect of this region on nitrogen/carbon balance before remobilisation. On chromosome 7B (wPt-3530-wPt-7113), Laperche et al. (2007) published a QTL of %N_S which colocalised with one of this study affecting the same trait. This region also appeared in Laperche et al. (2006) as being linked to the lateral root number and the primary root length, and in Habash et al. (2007) for GNC.

Breeding strategies

As we worked on a panel composed of commercial varieties mostly registered between 1985 and 2010, results of this study have to be discussed in light of selection pressures. Although QTL have been detected, if favourable alleles are already fixed in the more recent varieties, those QTL are not so useful in future breeding.

As expected, favourable alleles are more frequent in recent varieties for QTL affecting traits under a high selection pressure than on QTL affecting untargeted traits. We estimated a positive correlation ($P < 0.001$; $r^2 = 0.48$) between the frequencies of alleles having a positive effect (in varieties released from 2005) and genetic progresses assessed by Cormier et al. (2013). Cormier et al. (2013) showed that in this panel of European elite varieties, NUE was increased by improving N utilisation (NutE: +0.20 % year^{-1}) and remobilisation (NHI: +0.12 % year^{-1}; %N_S: -0.52 % year^{-1}) through a major positive selection pressure on grain yield (DMGY: +0.45 % year^{-1}), while maintaining constant N uptake. In agreement, we found that for DMGY QTL, NutE QTL, and %N_S QTL the median frequency of favourable alleles (in varieties released from 2005) were respectively 88, 68, and 79 % (Suppl. data 7). Moreover, for a given trait, the frequency of alleles having a positive effect in recent varieties is directly linked to the genetic correlation between this trait and DMGY ($P < 0.001$; $r^2 = 0.49$; Suppl. data 7). Thus, favourable alleles are already well represented in new varieties at QTL associated to traits directly (*e.g.* DMGY) or indirectly (*e.g.* NutE) targeted by breeding. This study has provided information to facilitate their monitoring.

Studying correlations between traits using QTL colocalisations rather than genetic correlations has the advantage of taking into account trait genetic architecture and the power with which we can

dissect them. Moreover, it gives a better estimation of the pleiotropic effect of QTL-based selection on a trait. Indeed, the genetic correlation is symmetric ($r_{a/b} = r_{b/a}$), contrary to the percentage of QTL colocalising between two traits. For example, based on our detection, selection on GPC QTL will surely affect NUE_Prot as all GPC QTL are also NUE_Prot QTL. However, only 73 % of QTL for GPC would be affected by selection on NUE_Prot QTL.

Results of colocalisation analyses revealed that we should select on INN_FLO, FLO, NutE, and %N_Flo QTL to maximise the number of affected traits. As 57 % (4/7) of INN_FLO QTL, and 50 % (4/8) of %N_Flo QTL were also FLO QTL, effect of phenology and pre-anthesis uptake are mixed. Thus, QTL controlling flowering time should be our first concern. Anthesis corresponds to a physiological transition and consequently, the date of this transition has a major impact on genotype × environment (G × E) interaction (Kamran et al. 2014). In this study, we observed an average genotypic flowering time standard deviation of 7 days. As varieties were tested in a small range of slightly contrasted environments, anthesis date directly affected G × E interaction and above all varieties' genetic values, favouring genotypes adapted to these environments. This created a confounding effect of major phenology genes (Reynolds et al. 2009) which are more likely to be associated to agronomic traits.

None of the central traits (INN_FLO, FLO, NutE, and %N_Flo; Fig 4) was linked to final N uptake. As mentioned before, recent breeding efforts improved N remobilisation and N utilisation, and not N uptake (Cormier et al. 2013). Thus, selection pressure enhanced N utilisation centrality in our network (Fig. 4). In this panel, the low genetic variance of the N uptake was not sufficient to reveal meaningful correlations with other agronomic traits and thus significant QTL colocalisations. Nevertheless, as a component of NUE, N uptake is a promising lever of action (Hirel et al. 2007; Foulkes et al. 2009). This study has provided tools to start selecting for N uptake in elite varieties without fastidious phenotyping or can be used as an entry point in investigating genes and pathways controlling this trait (Korte and Farlow, 2013) with further investigations in a more diverse panel.

Results on QTL colocalisations highlighted the importance of focusing on pre-anthesis nitrogen status, especially on INN_FLO which had a good heritability (0.63) and for which QTL have also the same effect on TKW and NUE_Prot.

CONCLUSIONS

Identification of chromosomal regions associated with nitrogen use efficiency-related traits at both high N levels and moderate N will help breeding for better adapted varieties. To our knowledge, this work is the first published study that reports GWAS results on N use efficiency in small grain cereals using a high marker density for precise mapping of genomic regions. Using an LD-based method to define QTL boundaries, 333 QTL were identified on 28 traits. Several colocalisations between our QTL and previously published QTL were pointed out. Using a network approach on colocalisation frequencies between traits, this study highlighted the interest of working on N status at flowering, and underscores the effect of recent breeding on N utilisation efficiency.

AUTHOR CONTRIBUTION STATEMENT

Statistical analyses and manuscript were conducted by FC during his PhD thesis co-directed by SP and JLG. PD provided useful help on power assessment methods and interpretations. JLG, SL, and SP were implicated in methods, interpretations, and in reviewing the manuscript.

ACKNOWLEDGEMENTS

These analyses were a part of a PhD thesis (129/2012) supported by the ANRT (Association Nationale de la Recherche et de la Technologie). Data were obtained thanks to the support of the ANR ProtNBle project (06 GPLA016).The authors are also grateful to experimental units staff at Estrées-Mons experimental unit (INRA) and at Villiers-le-Bâcle and Vraux (Arvalis), and to Quang Hien Le for English style proofing. Sincere thanks to Ian Mackay for his involvement in the reviewing process.

REFERENCES

An D, Su J, Liu Q, Zhu Y, Tong Y, Li J, Jing R, Li B, Li Z (2006). Mapping QTLs for nitrogen uptake in relation to the early growth of wheat (*Triticum aestivum* L.). Plant Soil 284:73-84

Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A, Joets J (2004) BioMercator: integrating genetic maps and QTN towards discovery of candidate genes. Bioinform 20:2324-2326

Avni R, Zhao R, Pearce S, Jun Y, Uauy C, Tabbita F, Fahima T, Slade A, Dubcovsky J, Distelfeld A (2013) Functional characterization of GPC-1 genes in hexaploid wheat. Planta 239:313-324

Barraclough PB, Howarth JR, Jones J, Lopez-Bellido R, Parmar S, Shepherd CE, Hawkesford MJ (2010) Nitrogen efficiency of wheat: genotypic and environmental variation and prospects for improvement. Eur J Agron 33:1-11

Bayer M, Milne I, Stephen G, Shaw P, Cardle L, Wright F, Marshall D (2011) Comparative visualization of genetic and physical maps with Strudel. Bioinformatics 27:1307-1308

Bordes J, Ravel C, Jaubertie JP, Duperrier B, Gardet O, Heumez E, Pissavy AL, Charmet G, Le Gouis J, Balfourrier F (2013) Genomic regions associated with the nitrogen limitation response revealed in a global wheat core collection. Theor Appl Genet 126:805-822

Breseghello F, Sorrells M (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165-1177

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) ASReml-R Reference Manual. Technical report, Queensland Department of Primary Industries. URL http://www.vsni.co.uk/software/asreml/

Cadalen T, Boeuf C, Bernard S, Bernard M (1997) An intervarietal molecular marker map in *Triticum aestivum* L. Em. Thell. and comparison with a map from a wide cross. Theor Appl Genet 94: 367-377

Conti V, Roncallo PF, Beaufort V, Cervigni GL, Miranda R, Jensen CA, Echenique VC (2011) Mapping of main and epistatic effect QTL associated to grain protein and gluten strength using a RIL population of durum wheat. J Appl Genet 52:287-298

Cormier F, Faure S, Dubreuil P, Heumez E, Beauchêne K, Lafarge S, Praud S, Le Gouis J (2013) A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (*Triticum eastivum* L.). Theor Appl Genet 126:3035–3048

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J, Reynolds M, Crouch JH, Ortiz R (2007) Association analysis of historical bread wheat germplasm using additive covariance of relatives and population structure. Genetics 177: 1889–1913

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Envir S 11:381-393

Distelfeld A, Avni R, Fischer AM (2014) Senescence, nutrient remobilization , and yield in wheat and barley. J Exp Bot 65:3783-3798

Dumur J, Jahier J, Bancel E, Laurière M, Bernard M, Branlard G (2004) Proteomic analysis of aneuploidy lines in the homeologous group 1 of the hexaploid wheat cultivar Courtot. Proteomics 4:2685-2695

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. Ed 4. Longmans Green, Harlow, Essex, UK

FAO (2012) World agriculture towards 203/2050, the 2012 revision.
http://www.fao.org/docrep/016/ap106e/ap106e.pdf

FAO (2011) Current world fertilizer trends and outlook to 2015.
ftp://ftp.fao.org/ag/agp/docs/cwfto15.pdf

Fontaine JX, Ravel C, Pageau K, Heumez E, Dubois F, Hirel, Le Gouis J (2009) A quantitative genetic study for elucidating the contribution of glutamine synthetase, glutamate dehydrogenase

and other nitrogen-related physiological traits to the agronomic performance of common wheat. Theor Appl Genet 119:645–662

Foulkes M, Hawkesford M, Barraclough P, Holdsworth M, Kerr S, Kightley S, Shewry P (2009) Identifying traits to improve the nitrogen economy of wheat: recent advances and future prospects. Field Crop Res 114:329–342

Gaju O, Allard V, Martre P, Snape JW, Heumez E, Le Gouis J, Moreau D, Bogard M, Griffiths S, Orford S, Hubbart S, Foulkes MJ (2011) Identification of traits to improve the nitrogen-use efficiency of wheat genotypes. Field Crop Res 123:139–152

Gregersen PL, Holm PB, Krupinska K (2008) Leaf senescence and nutrient remobilisation in barley and wheat. Plant Biol 10:37–49

Guarda G, Padovan S, Delogu G (2004) Grain yield, nitrogen-use efficiency and baking quality of old and modern Italian breadwheat cultivars grown at different nitrogen levels. Eur J Agron 21:181–192

Guo Y, Kong FM, Xu YF, Zhao Y, Liang X, Wang YY, An DG, Li SS (2012) QTL mapping for seedling traits in wheat grown under varying concentrations of N, P and K nutrients. Theor Appl Genet 124:851–865

Habash DZ, Bernard S, Schondelmaier J, Weyen J, Quarrie SA (2007) The genetics of nitrogen use in hexaploid wheat : N utilisation, development and yield. Theor Appl Genet 114:403–419

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theo Appl Genet 38:226-231

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54-78

Hirel B, Le Gouis J, Ney B, Gallais A (2007) The challenge of improving nitrogen use efficiency in crop plants: toward a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Bot 58:2369-2387

Hörtensteiner S, Feller U (2002) Nitrogen metabolism and remobilisation during senescence. J Exp Bot 53: 927-937

Huang A, Xu S, Cai X (2014) Whole-Genome Quantitative Trait Locus Mapping Reveals Major Role of Epistasis on Yield of Rice. PLoS ONE 9(1): e87330. doi:10.1371/journal.pone.008733

Justes E, Mary B, Meynard JM, Machet JM, Thelier-Huche L (1994) Determination of a critical nitrogen dilution curve for winter crops. Ann Bot-London 74: 397–407

Kamran A, Iqbal M, Spaner D (2014) Flowering time wheat (*Triticum aestivum* L.): a key factor for global adaptability. Euphytica 197:1-26

Kant S, Bi YM, Rothstein S (2011) Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. J Exp Bot 62:1499–1509

Knyazikhin Y, Schull M, Stenberg P, Mõttus M, Rautiainen M, Yang Y, Marshak A, Latorre Carmona P, Kaufmann R, Lewis P, Disney M, Vanderbilt V, Davis A, Baret F, Jacquemoud S, Lyapustin A, Myneni R (2013) Hyperspectral remote sensing of foliar nitrogen content. PNAS 110:185-192

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199

Laperche A, Devienne-Barret F, Maury O, Le Gouis J, Ney B (2006) A simplified conceptual model of carbon/nitrogen functioning for QTL analysis of wheat adaptation to nitrogen deficiency. Theor Appl Genet 113:1131-1146

Laperche A, Brancourt-Hulmel M, Heumez E, Gardet O, Hanocq E, Devienne-Barret F, Le Gouis J (2007). Using genotype x nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints. Theor Appl Genet 115:399–415

Larsen R, Marx M (1985) An introduction to probability and its applications. Englewood Cliffs, New Jersey; Prentice-Hall, Inc.

Le Gouis J, Beghin B, Heumez E, Pluchard P (2000) Genetic differences for nitrogen uptake and nitrogen utilisation efficiencies in winter wheat. Eur J Agron 12:163-173

Le Gouis J, Bordes J, Ravel C, Heumez E, Faure S, Praud S, Galic N, Remoué C, Balfourier F, Allard V, Rousset M (2012) Genome-wide association analysis to identify chromosomal regions determining components of earliness in wheat. Theor Appl Genet 124:597-611

Lorenz AJ, Hamblin MT, Jannink JL (2010) Performance of Single Nucleotide Polymorphisms versus Haplotypes for Genome-Wide Association Analysis in Barley. PLoS ONE 5(11): e14079. doi:10.1371/journal.pone.0014079

McCarthyMI, Hirschhorn JN (2008) Genome-wide association studies: potential next steps on a genetic journey. Hum Mol Genet 17:156-165

Mackay T (2014) Epistasis and quantitative traits: using model organisms to study gene–gene interactions. Nature Rev Genet 15:22-33

Mangin B, Goffinet B, Rebai A (1994) Constructing confidence intervals for QTL location. Genetics 138:1301-1308

Manske GGB, Ortiz-Monasterio IJ, Vlek PLG (2001) Techniques for measuring genetic diversity in roots. In: Application of physiology in wheat breeding. CIMMYT, Mexico, D.F, pp 208–218

Muurinen S, Slafer GA, Peltonen-Sainio P (2006) Breeding effects on nitrogen use efficiency of spring cereals under northern conditions. Crop Sci 46:561–568

Neumann K, Kobiljski B, Dencic S, Varshney RK, Börner A (2011) Genome-wide association mapping: A case study in bread wheat (Triticum aestivum L.). Mol Breed 27:37–58

Novoselovic D, Baric M, Drezner G, Gunjaca J, Lalic A (2004) Quantitative inheritance of some wheat plant traits. Genet Mol Biol 27:92-98

Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: Generalizing degree and shortest paths. Social Network 32:245-251

Ortiz-Monasterio I, Sayre KD, Rajaram S, McMahon M (1997) Genetic progress in wheat yield and nitrogen use efficiency under four N rates. Crop Sci 37:898–904

Pathak RR, Lochab S, Raghuram N (2011) Plant systems | improving plant nitrogen-use efficiency. In: Comprehensive biotechnology, 2nd edn. Elsevier, Amsterdam, pp 209–218

Payne P (1987) Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Am Rev Plant Physiol 38:141-153

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. Plos Genet 2:e190

Philipps PC (2008) Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet 9:855-867

Plessis A, Ravel C, Bordes J, Balfourier F, Martre P (2013) Association study of wheat grain protein composition reveals that gliadin and glutenin composition are trans-regulated by different chromosome regions. J Exp Bot 64:3627-3644

Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C, Steele N, Pljevljakusic D, Waterman E, Weyen J, Schondelmaier J, Habash DZ, Farmer P, Saker L, Clarkson DT, Abugalieva A, Yessimbekova M, Turuspekov Y, Abugalieva S, Tuberosa R, Sanguineti M-C, Hollington PA, Aragues R, Royo A, Dodig D (2005) A high density genetic map of hexaploid wheat (Triticum aestivum L.) from the cross Chinese Spring x SQ1 and its use to compare QTL for grain yield across a range of environments. Theor Appl Genet 110:865-880

Quraishi UM, Abrouk M, Murat F (2011) Cross-genome map based dissection of a nitrogen use efficiency ortho-metaQTL in bread wheat unravels concerted cereal genome evolution. The Plant Journal 65:745-756

Reynolds M, Manes Y, Izanloo A, Langridge P (2009) Phenotyping approaches for physiological breeding and gene discovery in wheat. An Appl Biol 155:309–320

Rincent R, Moreau L, Monod H, Kuhn E, Melchinger AE, Malvar RA, Moreno-Gonzalez J, Nicolas S, Madur D, Combes V, Dumas F, Altmann T, Brunel D, Ouzunova M, Flament P, Dubreuil P, Charcosset A, Mary-Huard T (2014). Recovering power in association mapping panels with variable levels of linkage disequilibrium. Genet 197:375-387

Rogers JS (1972) Measures of genetic similarity and genetic distances. Studies in Genetics, University of Texas Publication 7213:145-153

Rothstein S (2007) Returning to our roots: making plant biology research relevant to future challenges in agriculture. Plant Cell 19: 2695-2699

Somers D, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

Sourdille P, Gandon B, Chiquet V, Nicot N, Somers D, Murigneux A, Bernard M (2004) Wheat génoplante SSR mapping data release: a new set of markers and comprehensive genetic and physical mapping data. Funct Integr Genomics 4:12-25

Tester M, Langridge P (2010) Breeding Technologies to Increase Crop Production in a Changing World. Sci 327:818-822

Tian F, Bradbury PJ, Briwn PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, Mcmullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of maize identifies genes affecting leaf architecture. Nat Genet 43:159–162

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J (2006) A NAC gene regulating senescence improves grain protein, zinc and irons content in wheat. Science 314:1298–1300

Visscher P M, Thompson R, Haley CS (1996) Confidence intervals in QTL mapping by bootstrapping. Genetics 143:1013-1020

Wang S, Debbie W, Forrest K, Allen A, et al. (2014) Characterization of polyploidy wheat genomic diversity using a high-density 90 000 single

nucleotide polymorphism array. Plant Biotech J doi: 10.1111/pbi.12183

Wu X, Wang Z, Chang X, Jing R (2010) Genetic dissection of the developmental behaviours of plant height in wheat under diverse water regimes. J Exp Bot 61:2923-2937

Würschum T, Langer S, Longin CFH, Korzun V, Akhunov E, Ebmeyer E, Schachschneider R, Kazman E, Schacht J, Reif JC (2013) Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor Appl Genet 126:1477-1486

Xu Y, Wang R, Tong Y, Zhao H, Xie Q, Liu D, Zhang A, Li B, Xu H, An D (2013) Mapping QTL for yield and nitrogen-related traits in wheat: influence of nitrogen and phosphorus fertilization on QTL expression. Theor Appl Genet 127:59-72

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang K, Tian J, Zhao L, Wang S (2008) Mapping QTL with epistatic effects and QTL x environment interactions for plant height using a doubled haploid population in cultivated wheat. J Genet Genomics 35:119-127

Zhang LY, Liu DC, Guo XL, Yang WL, Sun JZ, Wang DW, Zhang A (2010) Genomic distribution of quantitative trait loci for yield and yield-related traits in common wheat. J Integr Plant Biol 52:996-1007

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, et al. (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat Comm 2:467-477.

AN EXAMPLE OF CANDIDATE GENE DISCOVERY: *NAM-A1*

Applying an empiric method to define quantitative trait locus (QTL) from results of genome-wide association study (GWAS), we found 333 QTL for 28 traits. QTL mean size was relatively small (3.2 cM). Thus, we concluded that GWAS-based QTL cloning can be a good alternative and speed up the classical QTL cloning approach. Nevertheless, we should keep in mind that QTL size variability was high. Indeed, 90% of our QTL had a size between 0.1 cM and 14 cM (5^{th} and 95^{th} percentiles). Using the recent estimation of gene density in wheat [1] we estimated that these QTLs contain between 1 and 2,000 genes. Therefore, QTL selection and data mining to screen candidate genes are essential. To illustrate this, here, we will detail the work based on a QTL that actually appeared in the previous paper as GNY5 (see Annexes of Part III) and where we highlighted the importance of the most interesting candidate gene named *NAM-A1*. Characterization of *NAM-A1* natural variants was submitted for publication to Agronomy.

NAM-A1 a good candidate gene

GNY5 is a small QTL (0.64 cM) of grain nitrogen yield located on chromosome 6A around 56.5 cM in Biogemma genetic map. Previously to this PhD thesis, GWAS conducted in Biogemma identified this region as associated with yield related traits. Multi-environmental GWAS performed during this PhD thesis (not presented in this manuscript) also revealed that this region had an effect on nitrogen use efficiency (NUE) that significantly interacted with the level of applied nitrogen (N). Added to that, this region is homeologous of the *Gpc-B1* locus (Fig. 5).

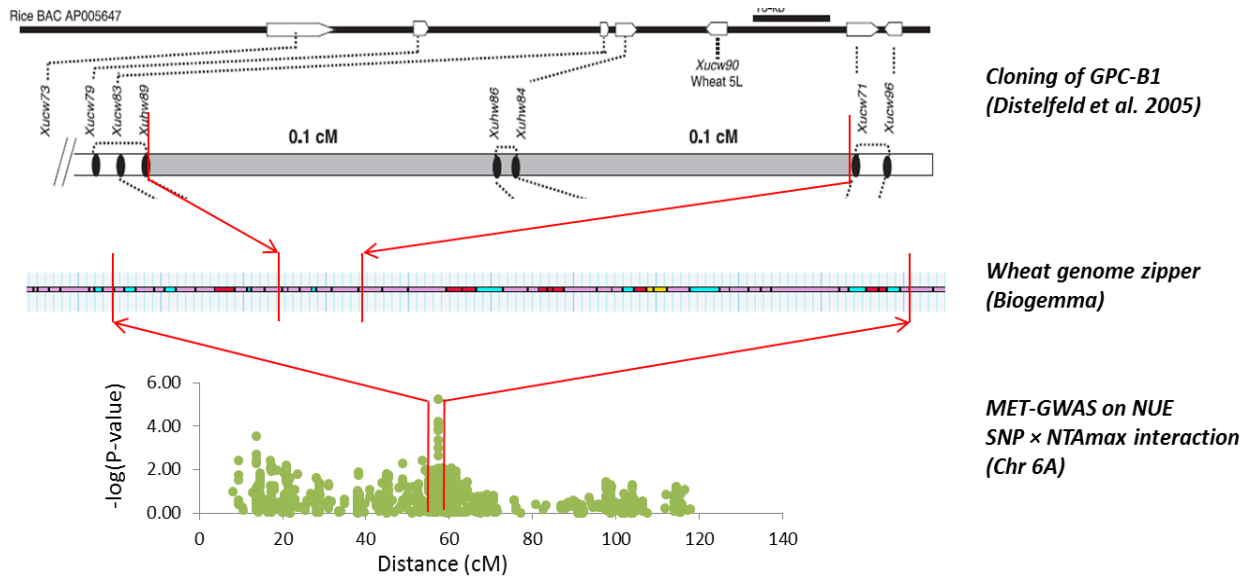


Figure 5: GNY5 and Gpc-B1 are homeologous. The MET-GWAS model was a mixed model K including quality and precocity as covariates, a SNP main effect and an SNP × NTA_{max} interaction.

In hexaploid bread wheat (*Triticum aestivum* L.) and tetraploid durum wheat (*Triticum turgidum* L. ssp durum) the No Apical Meristem (NAM) gene at the *Gpc-B1* locus (*NAM-B1*) on chromosome arm 6BS encodes a NAC transcriptional factor known to accelerate senescence and to increase nutrient remobilisation [2-4] hence grain protein concentration. Different effects of *NAM-B1* were assessed depending on genotypes × environment combinations [2]. Moreover, optimal senescence kinetic can differ depending on N levels [5] leading to the hypothesis that *NAM-B1* effects can also depend on the fertilisation regimes.

Most bread wheats have a non-functional allele of *NAM-B1* [6]. Consequently, its physiological characterization began after a chromosome segment introgression from wild emmer wheat (*Triticum turgidum* L. subsp. dicoccoides) [7]. Nevertheless, hexaploid wheats have five other NAM genes, two homoeologous (on chromosomes 6A and 6D) and three paralogous (on chromosomes 2A, 2B and 2D) of which *NAM-A1* (6A) has the same role as *NAM-B1* [4,8]. Consequently, *NAM-A1* was a good candidate for the GNY5 QTL. Most studies on NAM wheat genes used mutants [4, 8], near isogenic lines [9-12] or RNAi lines [3,4] and few studies focused on the cultivated diversity [6,13]. Thus, we aimed to characterize natural variants of *NAM-A1* in hexaploid bread wheat and to hypothesize biological mechanism involved in their putative effects to validate this gene as a good candidate.

SNP detection

We screened the IWGSC (International Wheat Genome Sequencing Consortium) bank of genomic sequences and identified *NAM-A1* in the sequence 6AS:4397602. In this 29,595 pb sequence composed of several transposable elements, the coding sequence of *NAM-A1* was localized between 15,502 bp and 17,060 pb and is composed of three exons for a total length cDNA length of 1,235 pb.

SNP (single nucleotide polymorphism) identification was performed on 12 varieties and two high quality SNP were detected in *NAM-A1* genomic region (Suppl. data 8). The first SNP (SNP1) was located in *NAM-A1* NAC domain (exon 2, 6AS:4397602_16233) and tagged a C/T polymorphism. This SNP caused an alanine to valine substitution in the protein sequence. The second SNP (SNP2) was located at the end of the coding sequence (exon 3, 6AS:4397602_17020) tagged an A/deletion polymorphism and caused a reading frame shift leading to a truncated protein (Suppl. data 9).

Using the KASPar technology, these two SNP were genotyped on a total of 795 wheat cultivars composed of the 367 worldwide core collection accessions [14] and 334 elite varieties with six varieties in common. Computing linkage disequilibrium between SNP located in *NAM-A1* and SNP from the iSelect 90K wheat SNP chip [15], we confirmed that our SNP tagging *NAM-A1* were located on chromosome 6A in GNY5.

SNP frequencies were not balanced (Table 5). For SNP1, the T allele was the most frequent in the core and elite collections (0.747 and 0.915 respectively). For SNP2, the A allele was more frequent in the core collection (0.765) and the Del allele in the elite collection (0.724). When considering haplotypes, *NAM-A1c* (T-A) was the most frequent haplotype in the core collection and *NAM-A1d* (T-Del) in the elite panel. In the core collection, accessions carrying the haplotype *NAM-A1d* were then mainly Western European

modern cultivars released intentionally added in the core collection. In both panels, haplotype *NAM-A1b* was the less frequent with no accession carrying it in the elite panel and only one landrace from Georgia in the core collection. A Khi² test shows that the observed haplotypes frequencies are not as expected from the SNP frequencies (Khi² = 120, P < 0.001, both collections together, Suppl. data 10). Although *NAM-A1d* is not the major haplotype in the core collection, it is over-represented in the two collections together. The *NAM-A1a* haplotype is also over-represented while the *NAM-A1b* is largely under-represented.

Table 5: NAM-A1 haplotype frequencies on two collections of bread wheat genotypes. Frequency followed by the number of lines (in parenthesis).

Genotype			Frequency	
SNP1	**SNP2**	**Haplotype**	**Core Collection**	**Elite**
C	A	*NAM-A1a*	0.232 (85)	0.083 (28)
C	Del	*NAM-A1b*	0.003 (1)	0.000 (0)
T	A	*NAM-A1c*	0.477 (175)	0.189 (63)
T	Del	*NAM-A1d*	0.215 (79)	0.716 (239)
		Undefined	0.074 (27)	0.012 (4)

In the worldwide core collection, NAM-A1a is mainly found in accessions from Nepal (23 of 21), China (16 of 8) and Japan (12 of 7). Moreover, accessions carrying the haplotype *NAM-A1a* are mostly spring wheat. In the elite collection, *NAM-A1a* is over-represented in varieties with a high bread-making quality. Brevis et al. [10] showed that Gpc-B1 introgression was associated with a positive effect on several bread-making and pasta-making quality parameters. We can expect the same effect for *NAM-A1*. Thus, *NAM-A1a* may have been maintained in elite germplasm through selection for high baking quality. Added to that, SNP1 is linked to the core collection genetic structure as SNP1_C is over-represented in far Eastern countries that form a cluster of diversity in the core collection [14]. Consequently, *NAM-A1b* under-representation could probably be explained by a Del mutation (SNP2) occurring only in the SNP1_T allelic lineage [16]. Then, over-representation of *NAM-A1d* in modern European elites suggests that the haplotype may have been selected. *NAM-A1b* could be the results of a recent recombination between *NAM-A1a* and *NAM-A1d*.

Effect of *NAM-A1* haplotypes

Focusing on the 196 European elite varieties genotyped in this study and belonging to the phenotyping dataset used in this PhD thesis [17], effects of *NAM-A1* haplotypes were the most significant effects detected in the *NAM-A1* chromosomal region (Suppl. data 11). The highest grain protein concentration (GPC) and lowest grain yield (GY) were reached in varieties carrying the haplotype *NAM-A1a* (Table 2). This is caused by the well-known negative correlation between GY and GPC (*i.e.* [18]). The lower grain yield was linked with a reduced grain weight (TKW) not compensated by the number of grain [spike per

area (SA) × kernel per spike (KS) in Table 6]. Nevertheless, varieties with *NAM-A1a* showed also the highest grain protein deviation (GPD, [19]) and a high N harvest index associated with a low straw N content at maturity (%N_S). Varieties carrying the haplotype *NAM-A1c* were intermediate between those carrying *NAM-A1a* and *NAM-A1d*. This can be explained by differences in haplotype effects. However, varieties genetic background effect is also a possible explanation. In general, due to the highly unbalanced frequencies and a distribution linked to the panels structure as previously mentioned, we lacked power to be able to distinguish the effect of genotypes genetic background and the actual effect of *NAM-A1*.

Table 6: Mean agronomic values for the two *NAM-A1* SNP genotyped on 196 (16 *NAM-A1a*; 37 *NAM-A1c*; 143 *NAM-A1d*) European elite varieties.

SNP1	SNP2	Haplotype	GY	TKW	SA	KS	GPC	GPD	NHI	%N_S
C	A	*NAM-A1a*	6,976c	41.3b	421a	40.4b	10.46a	0.20a	81.17ab	0.41a
T	A	*NAM-A1c*	7,241b	41.6b	413a	42.5a	10.15b	0.04ab	81.47a	0.41a
T	Del	*NAM-A1d*	7,799a	42.7a	411a	43.0a	9.79c	-0.09b	80.98b	0.42b

GY, dry matter grain yield (kg/ha); TKW, thousand kernel weight (g); SA, spike per area (spike/m²); KS, kernel per spike; GPC, Grain Protein Concentration (%) ; GPD, Grain Protein Deviation [19]; NHI, nitrogen harvest index (%N); %N_S, straw N content at maturity (%N). Letters indicate significance group by LSD test (P<0.05).

Nevertheless, in agreement with the described mean values, several studies analyzing the introgression of the functional allele of *Gpc-B1* in different spring hexaploid wheat [9, 11, 12] concluded that *NAM-A1* homoeolog increased GPC and decreased TKW. An improved N remobilisation (%N_S and NHI) was also assessed [9]. However, the effect of *Gpc-B1* on grain yield across genotypes and environments was not significant [9, 11, 12] even if it was strongly affected by the genetic background [9]. In the same way, study of mutants concluded that functional *NAM-A1* (6A) and *NAM-B2* (2B) genes accelerate senescence and increase GPC with a larger phenotypic effect for *NAM-A1* than *NAM-B2* [4, 8].

To conclude, we hypothesized that *NAM-A1a* could be a functional variant of *NAM-A1* gene. Accelerated senescence could have improved N remobilisation and GPC but decreased TKW leading to a GY decrease as in our elite panel where varieties carrying *NAM-A1a* had also a lower number of grains and/or are more likely to benefit from a stay-green phenotype in the tested environment. This is in accordance with the low frequency of *NAM-A1a* in elite germplasm mainly selected on GY, and its high frequency in spring Nepalese accessions cultivated within a short growing season.

Prediction of 3D structure

Prediction of the NAM-A1 NAC domain 3D structure was based on the crystal structure of the rice stress responsive NAC1 (SNAC1) NAC domain [20]. Crystallographic analysis of the NAC domain of the ANAC protein [21, 22] encoded by the abscisic acid-responsive NAC gene from Arabidopsis thaliana and mutants study [23] were also used.

According to their high amino acid similarity (69.7%), the topology of SNAC1 NAC domain and the predicted topology of NAM-A1 NAC domain were similar. The NAM-A1 NAC domain prediction resulted in seven twisted β-strands forming a semi-β-barrel with four α-helices (Fig. 6). Although, the residues of the loop region between β6–β7 in both SNAC1 and ANAC NAC domains were unobserved due to its non-participation in crystal packing [20], in NAM-A1 NAC domain an α-helix is predicted. This α4-helix is truncated in the protein encoded by the haplotypes *NAM-A1c* and *NAM-A1d*, due to SNP1 alanine to valine substitution (Fig. 6). Indeed, alanine is one of the best α-helix-forming residues due to aliphatic sidechains regions. At the opposite, with short sidechains that can form hydrogen bonds, valine is a poor α-helix former.

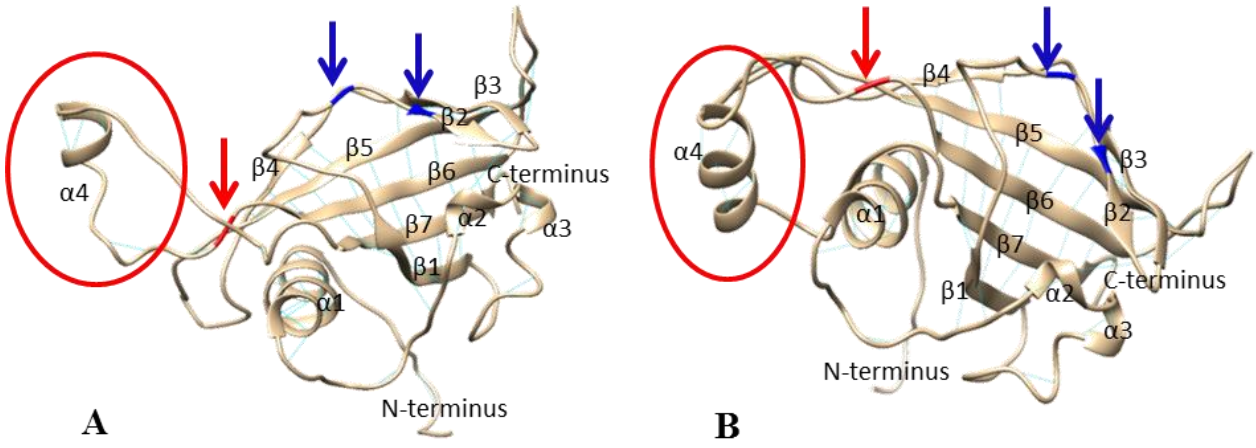


Figure 6: Predicted 3D structure of NAM-A1 NAC domain for (A) the valine variant (SNP1_T) and (B) the alanine variant (SNP1_C). Blue arrows: Arg107 and Arg110; red arrows: variant amino acid; red circle: affected α4-helix.

Dimerization of DNA binding domains is common and can modulate the DNA-binding specificity [24]. Gel filtration studies on ANAC NAC domain [21] and SNAC1 NAC domain [20] have shown that in solution they exist as dimers that form the functional unit necessary for stable DNA binding [23]. We can reasonably presume it is also the case for NAM-A1. The interface between the two monomers of SNAC1 consists of residues in the N-terminal loop region and two residues in the α1-helix [20]. In NAM-A1, this domain is not predicted to be affected by SNP1 variation.

Olsen et al. [23] showed that K79A/R85A/R88A and R85A/R88A were ANAC mutants that impaired DNA binding. Using these results, Chen et al. [20] hypothesized that Arg85 and Arg88 were responsible of DNA binding in SNAC1 (residues Arg107 and Arg110 in NAM-A1).

Using yeast one hybrid assay, Duval et al. [25] identified the DNA binding domain of AtNAM between Val119 and Ser183 (AtNAM numbering) and hypothesized that the region folds in a helix-turn-helix structure. In contrast, in ANAC and SNAC1, this region consists of β-sheet [20, 22], but as previously mentioned the conformation of part of residues in the loop region between β6–β7 was unobserved. This unobserved loop region poorly conserved between NAC domains and maybe related to their biologic

function [20] was predicted as the region affected by the alanine to valine substitution discovered in *NAM-A1*.

Thus, in accordance with the lowest GPC and GPD observed (Table 2) for the *NAM-A1d* (SNP1_T, SNP2_del) haplotype compared to the *NAM-A1a* haplotype (SNP1_C, SNP2_A), we hypothesize that the valine variant of NAM-A1 NAC domain (SNP1_T) may form dimers, bind to DNA, but its biological function is affected. A second hypothesis could be that the more recent mutation (SNP2) leading to a slightly truncated protein may affect the transcriptional activation by the C-terminus and difference between *NAM-A1a* and *NAM-A1c* could be due to genetic background effect. Sequence alignment of closest NAC proteins from wheat, barley, rice and *A. thaliana* did not allow comparing the two hypothesis as these NAC proteins mostly carry the alanine variant and none of them seems truncated (Suppl. data 12).

Conclusion on *NAM-A1*

Grain protein concentration was maximized in varieties carrying the *NAM-A1a* haplotype coding for the alanine variant of *NAM-A1* NAC domain and a non-truncated protein confirming the hypothesis that it may be a functional haplotype conserved in high-baking quality germplasm used in modern selection. Understanding the difference between both haplotypes coding a valine variant of *NAM-A1* NAC domain (*NAM-A1c* and *NAM-A1d*) remained unclear. Thus, further investigation at low N regime after flowering may be required to maximize the impact of remobilisation on agronomic performance. In the context of fertiliser reduction, increasing the frequency of the *NAM-A1a* haplotype in elite germplasm may help to breed for an increased remobilisation. Effect of *NAM-A1* on yield seemed to depend on genotypes and environments. This study provided the tools for further investigations.

The example of *NAM-A1* illustrates the interest in confronting different sources of information to finally end with a candidate gene. Moreover, using multi-environmental data helps (i) to build a hypothesis on the biological mechanisms involved and (ii) to design future experiments. Nevertheless, for quantitative traits, implementation of the knowledge resulting from this approach can be limited in breeding programs. Indeed, even if QTL cloning is sped up, we work on small effect loci hoping that their combine responses will be additive. However, the demand of varieties with an enhanced NUE is urging and genomic selection approaches may appear more attractive to breeders. Nevertheless, could we use MET-GWAS results to increase GS efficiency?

Contribution

Mickael Throude: reconstruction of GNY5 chromosomal region, SNP detection, coordination of genotyping, interpretation. Nadine Duranton and Florence Exbrayat: collection genotyping. Catherine Ravel: SNP development, core collection genotyping, interpretation. Magalie Leveugle: sequencing, bioinformatics. Fabien Cormier: elite collection genotyping, SNP detection, 3D conformation, interpretation. Jacques Le Gouis, Stéphane Lafarge and Sébastien Praud: supervision and interpretation.

Materials and methods

Results of the MET-GWAS plotted in Fig. 1 were obtained with the following mixed model:

$$y_{ij} = \mu + e_j + q_i + bf_i + \alpha_i + \beta_i NTA_{max} + u_i + \varepsilon_{ij}$$

where y_{ij} is the phenotypic value of genotype i environment j, μ is the trait general mean, ej the effect of j, qi the effect of the quality class of i, b the general sensitivity to flowering time, f_i the mean flowering date of i, α_i the allele of genotype i at marker α, βi the sensitivity of allele α_i to the NTA_{max}, $ui \sim N(0, \sigma_u^2 \mathbf{K})$ a genetic background effect with **K** a matrix of relative kinship, and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ a residual error term.

The IWGSC (International Wheat Genome Sequencing Consortium) bank of genomic sequences was screened by Basic Local Alignment Search Tool (BLAST) using the sequence DQ869672.1 (*Triticum turgidum* subsp. durum *NAM-A1* complete coding DNA sequence).

SNP detection was performed following sequencing of *NAM-A1* in 12 varieties: Alcedo, Brigadier, Cassius, Premio, Récital, Renan, Rialto, Robigus, Sarina, Soissons, Tremie and Xi19. Genomic sequences were aligned using Chinese Spring as a reference.

The KASPar SNP Genotyping System (KBiociences, Herts,UK) was used to validate SNPs. KASPar Primers were designed with Primer picker (KBioscience) and PCR amplifications were performed on hydrocycler (LGC genomics), for 50 cycles at 57°C and then run onto a Genotyper (Applied Biosystem).

Linkage disequilibrium between the discovered SNP on *NAM-A1* and the iSelect 90K SNP was computed using genotyping data of 281 varieties from the European elite collection.

Mean agronomic values were calculated from 196 European elite varieties (16 CA; 37 TA; 143 TDel) experimented in eight combinations of year, site, and N regime [17]. Mean values were calculated using a linear model with the experiment (year_site_N) and SNP or haplotype as fixed factors.

Prediction of 3D structure was carried out using SWISS-MODEL SERVER [26] and based on the 3ulx.1.A template (X-ray, 2.60 Å) of SNAC1 [20]. Visualization was made using Chimera [27].

References

1. Choulet F, Alberti A, Theil S, Glover N, Barbe V, et al. (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science DOI: 10.1126/science.1249721
2. Uauy, C., Brevis J.C., Dubcovsky J. The high grain protein content gene Gpc-B1 accelerates senescence and has pleiotropic effects on protein content in wheat. J Exp Bot 2006, 57, 2785–2794.
3. Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., Dubcovsky, J. A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 2006, 1298–1301.
4. Distelfeld, A., Pearce S. P., Avni R., Scherer B., Uauy C., Piston F., et al. Divergent functions of orthologous NAC transcription factors in wheat and rice. Plant Mol. Biol. 2012, 78, 515–524.
5. Gaju, O., Allard, V., Martre, P., Snape, J.W., Heumez, E., Le Gouis, J., Moreau, D., Bogard, M., Griffiths, S., Orford, S., Hubbart, S., Foulkes, M.J. Identification of traits to improve the nitrogen-use efficiency of wheat genotypes. Field Crops Res 2011, 123, 139-152.
6. Hagenblad, J., Asplund, L., Balfourier, F., Ravel, C., Leino, M. Strong presence of the high grain protein content allele of NAM-B1 in Fennoscandian wheat. Theor Appl Genet 2012, 125, 1677–1686.
7. Distefeld, A., Uauy, C., Fahima, T., Dubcosky, J. Physical map of the wheat high-grain protein content gene Gpc-B1 and development of a high-throughput molecular marker. New Physiologist 2006, 169, 753-763.

8. Avni, R., Zhao, R., Pearce, S., Jun, Y., Uauy, C., Tabbita, F., Fahima, T., Slade, A., Dubcovsky, J., Distelfeld, A. Functional characterization of GPC-1 genes in hexaploid wheat. Planta 2014, 239, 313-324.
9. Brevis, J.C., Dubcovsky, J. Effects of the Chromosome Region Including the Gpc-B1 Locus on Wheat Grain and Protein Yield. Crop Sci 2010, 50, 93–104.
10. Brevis, J.C., Morris, C.F., Manthey, F., Dubcovsky, J. Effect of the grain protein content locus Gpc-B1 on bread and pasta quality. J Cereal Sci 2010, 51, 357-365.
11. Eagles, H.A., McLean, R., Eastwood, R.F., Appelbee, M.J., Cane, K., Martin, P.J., Wallwork, H. High-yielding lines of wheat carrying Gpc-B1 adapted to Mediterranean-type environments of the south and west of Australia. Crop & Pasture Science 2014, 65, 854-861.
12. Tabbita, F., Lewis, S., Vouilloz, J.P., Ortega, M.A., Kade, M., Abbate P.E., Barneix, A.J. Effects of the Gpc-B1 locus on high grain protein content introgressed into Argentinean wheat germplasm. Plant Breeding 2012, 132, 48-52.
13. Asplund, L.,Bergkvist, G., Leino, M.W., Westerbergh, A., Weih, M. Swedish spring varieties with the rare high grain protein allele of NAM-B1 differ leaf senescence and grain mineral content. PloS One 2013, 8(3): e59704.
14. Balfourier, F., Roussel, V., Strelchenko, P., Exbrayat-Vinson, F., Sourdille, P., et al. A worldwide bread wheat core collection arrayed in a 384-well plate. Theor Appl Genet 2007, 114, 1265–1275.
15. Wang, S., Debbie, W., Forrest, K., Allen, A., et al. Characterization of polyploidy wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotech J 2014, 12, 787-796.
16. Flint-Garcia, S., Thornberry, J.M., Buckler, E.S. Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 2003, 54, 357-374.
17. Cormier, F., Faure, S., Dubreuil, P., Heumez, E., Beauchêne, K., Lafarge, S., Praud, S., Le Gouis, J. A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (Triticum aestivum L.). Theor Appl Genet 2013, 126, 3035–3048.
18. Feil, B. The inverse yield-protein relationships in cereals: possibilities and limitations for genetically improving the grain protein yield. Trends in Agronomy 1997, 1, 103-119.
19. Monaghan, J.M., Snape, J.W., Chojecki, A.J.S., Kettlewell, P.S. The use of grain protein deviation for identifying wheat cultivars with high protein concentration and yield. Euphytica 2001, 122, 309–317.
20. Chen, Q., Wang, Q., Xiong, L., Lou, Z. A structural view of the conserved domain of rice stress-responsive NAC1. Protein Cell 2011, 2:55-63.
21. Olsen, A.N., Ernst, H.A., Lo Leggio, L., Johansson, E., Larsen, S., Skriver, K. Preliminary crystallographic analysis of the NAC domain of ANAC, a member of the plant-specific NAC transcription factor family. Acta Crystallogr D 2004, 60, 112–115.
22. Ernst, H.A., Olsen, A.N., Skriver, K., Larsen, S., Lo Leggio, L. Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. EMBO Rep. 2004, 3, 297-303.
23. Olsen, A.N., Ernst, H.A., Leggio, L.L., Skriver, K. NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci. 2005, 10, 79-87.
24. Müller, C.W. Transcription factors: global and detailed views. Curr Opin Struct Biol 2001, 11, 26-32.
25. Duval, M., Hsieh, T.F., Kim, S.Y., Thomas, T.L. Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. Plant Mol Bio 2002, 50, 237-248.
26. Arnold, K., Bordoli, L., Kopp, J., Schwede, T. The SWISS-MODEL workspace: a web based environment for protein structure homology modelling. Bioinformatics 2006, 22, 195-201.
27. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M. et al. UCSF Chimera a visualization system for exploratory research and analysis. J Comput Chem. 2004, 25, 1605-12.

PART IV: ALL TOGETHER NOW

This work was hosted by the CIMMYT
This article is ready to be submitted
Estimated submission: April 2015

IMPROVING GENOMIC PREDICTION USING A GWAS-BASED METHOD TO PRE-SELECT MARKERS IN MULTI-ENVIRONMENT TRIALS

Fabien Cormier[1] ✉ • **Delphine Ly**[2] • **Jacques Le Gouis** [2] • **Sébastien Praud**[1] • **Jose Crossa**[3]

[1]Biogemma, Centre de Recherche de Chappes, Chappes, France. [2]UMR 1095 Génétique, Diversité et Ecophysiologie des Céréales, INRA - Université Blaise Pascal, Clermont-Ferrand, France. [3]Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), El Batan, Mexico
✉ fabien.cormier@biogemma.com

ABSTRACT: Recently, the development of genome-wide prediction methods has experienced a burst exploring a broad diversity of approaches. Nevertheless, the widespread assumption that no specific knowledge of causal loci is required may have to be reconsidered. Moreover, prediction of genotype-by-environment interaction remains a major issue. We performed a multi-environment genome-wide association study (MET-GWAS) including marker-by-environmental covariate interactions to rank markers by significance of their main effect and significance of their interaction with environmental covariates. We used these rankings and the number of markers as two independent parameters and assessed genomic prediction accuracies in three cross-validation designs. In this study, we concluded that genomic prediction efficiency can be easily increased using marker pre-selection based on MET-GWAS results. Depending on the studied trait, we reduced the number of markers used from 25,368 to 1,275 and 700, and we increased the prediction accuracy of new genotypes from 0.52 and 0.25 to 0.61 and 0.44, respectively. For prediction in incomplete designs or for new environments, we drastically reduced the number of markers and maintained high prediction accuracy. We showed that reducing the number of markers for genetic value prediction increased accuracy stability. Depending on the cross-validation design, genotype-by-environment variance from 17.6 % to 30.2 % was predicted using markers and simple environmental characterization. This study is a first step toward using preliminary knowledge of genetic architecture in multi-environment genomic prediction.

INTRODUCTION

Genome-wide association studies (GWAS) and genomic predictions are often considered as two different approaches used to achieve different objectives. GWAS which assesses loci effects independently from each other is mainly used to discover genes or in genetic architecture studies assuming that traits are controlled by a relatively small number of quantitative trait loci (QTLs). Genomic prediction hypothesizes that a large number of loci in the genome have an effect on complex traits, and takes into account even the smallest effects that dominate complex traits to predict genotypes' performance.

The increased number of markers available thanks to the development of high-throughput genotyping methods has made GWAS results more and more difficult to implement in routine marker-assisted selection [1]. Moreover, loci effects are clearly misestimated in GWAS, and confounding due to genetic relatedness remains a major trade-off [2]. In the meantime, genomic prediction methods experienced a burst and appear promising in breeding strategies [3-5].

Several studies have highlighted the impact of the number of causal loci on the accuracy of genomic prediction (for example [6]). Similarly, various studies have assessed the effect on accuracy of the number of markers used in genomic prediction of various traits in animal or plant species [7-11]. Their results have led to the conclusion that the common assumption that no specific knowledge of causal loci location is required for genomic prediction might have to be re-considered. Thus, when marker pre-selection is needed, two problems arise: (i) the number of markers pre-selected and (ii) the criteria used to select them. Different methods that reduce the number of markers have been tested such as pre-selection based on a previous step of marker effect estimation [7-9] or the use of GWAS results [10, 12]. Nevertheless, the number of markers used was always increased by adding marker from the first to the last, making it impossible to independently assess the effect of the number of markers and marker rank. Moreover, these studies focused on genetic value and did not address the issue of genotype-by-environment (G × E) interaction prediction.

More generally, to date, genomic prediction methods focus mostly on predicting genetic values of complex traits. However, in plant breeding, G × E interactions remain a major limitation, as they can contribute significantly to genetic variance that leads to changes in ranking between environments [13]. This complicates selection for broad adaptation, especially in the context of climate change and inputs reductions which inflate G × E contributions. Genotype-by-environment interaction was first introduced in genomic prediction models using structured covariance between environments [14]. Then, to be able to predict genotypes response to new environments, environmental covariates (ECs) were introduced using factorial regression [15] or a reaction norm framework model [16]. Numerous ECs can be derived from environmental factors such as temperature or rainfall [16] or crop model [15] leading to variable selection issues.

This study aims to suggest a method for increasing genomic prediction efficiency using GWAS results for both genetic values and G × E interactions. The strategy described here is based on single-nucleotide polymorphism (SNP) pre-selection and was designed to be easy to implement. In view of

future agriculture challenges, and societal and environmental concerns, we chose to work on complex traits related to nitrogen use (N): nitrogen use efficiency (NUE) and nitrogen harvest index (NHI). Traditional phenotyping methods for NUE and NHI are labor intensive and partially destructive. Thus, they cannot be easily implemented in breeding programs and require marker-assisted selection. Regarding their polygenic genetic determinism, genomic selection is one of the best options.

RESULTS

Variance analysis

We studied two traits related to nitrogen use (NUE and NHI) in wheat using a 214-variety panel evaluated in eight environments that are defined as a year × location × N combination. For both traits G × E interactions were significant ($P < 0.001$) and explained 23 % of the within environment variance for NUE and 16 % for NHI (Table 1). Residuals were high and accounted for 29 % of the variance for NUE and 69 % for NHI. Nevertheless, the generalized heritabilities of NUE and NHI were 0.88 and 0.62, respectively, given that we worked on data resulting from precise phenotyping assessed in several within environment replications. Regarding variance decomposition, if we had succeeded in predicting all genetic (G + G × E) variance, the maximum accuracy for prediction of phenotypic values would have been about 0.84 for NUE and 0.56 for NHI.

Effect of SNP number and rank on prediction of additive genetic values

To evaluate the effect of the number of markers and marker significance in MET-GWAS on genomic prediction of both NUE and NHI traits, SNP were ranked according to their significance in multi-environment genome-wide association studies (MET-GWAS). In these rankings, we defined different SNP sections of significance. The number of SNPs (section size) and the section rank used in our genomic prediction model (an extension of G-BLUP) were two independent parameters. To avoid redundancy, we used a total of 2,101 SNPs that we pre-selected based on linkage disequilibrium (LD).
First, we studied the correlation ($r_{(Gi/gi)}$) between genetic value (G_i) and its predictor (g_i). The major prediction issue for genetic value occurred in cross-validation 1 (CV1) as one-third of the varieties had never been evaluated in any environment.

Table 1: Estimation of variance components. Percentage relative to the total within environment variance. Estimated values are in brackets.

Trait	H^2g	Model	σ^2_G	σ^2_{GE}	σ^2_ε	r_{max}
NUE	0.88	$G_i + \varepsilon_{ik}$	51 (8.14)		49 (7.84)	
	0.88	$G_i + GE_{ij} + \varepsilon_{ijk}$	48 (7.74)	23 (3.71)	29 (4.72)	0.84
NHI	0.62	$G_i + \varepsilon_{ik}$	17 (1.76)		83 (8.64)	
	0.62	$G_i + GE_{ij} + \varepsilon_{ijk}$	15 (1.52)	16 (1.71)	69 (7.17)	0.56

$r_{max} = \sqrt{\left(\sigma^2_G + \sigma^2_{GE}\right) / \sigma^2_\varepsilon}$ *is the theoretical maximum accuracy for phenotypic value prediction.*

G: genotype; GE: genotype × environment; ε: model residual.

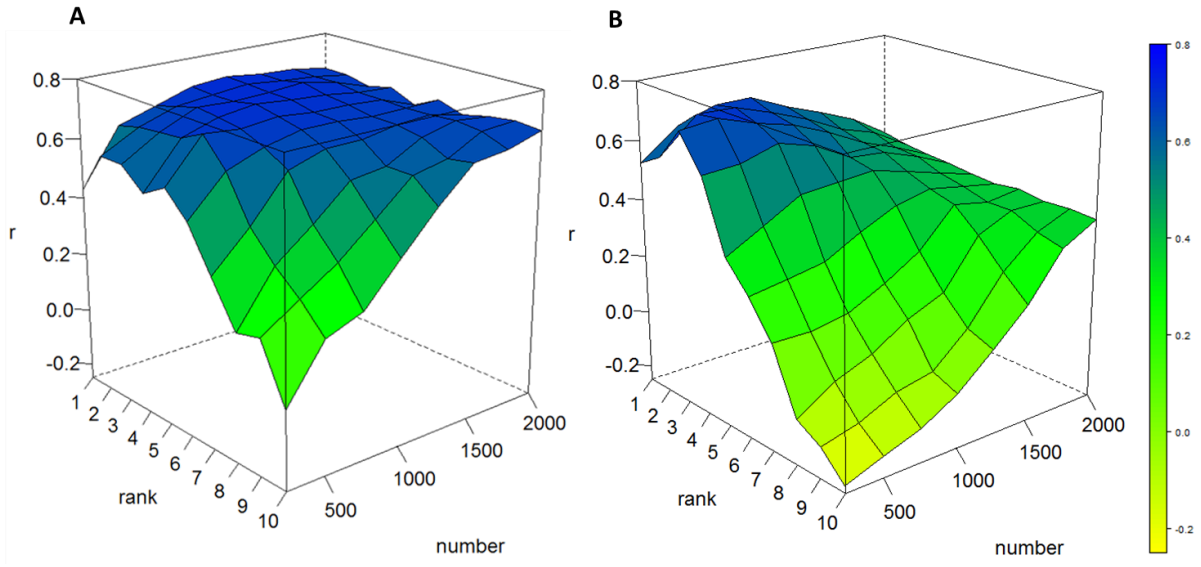


Figure 1: Evolution of genetic value prediction accuracy of (A) NUE and (B) NHI. Predictions were assessed using a three-fold design repeated 50 times for each combination of SNP number and SNP section rank in GWAS-based SNP ranking.

In CV1, prediction accuracy of NUE genetic values was highest when the 1,250 most significant SNPs were used (r = 0.78 +/- 0.21; Fig. 1A). For NHI, prediction accuracy was highest (r = 0.70 +/- 0.20) when the 500 SNPs of the third section were used (Fig. 1B). For each trait, this optimal combination of section size and section rank also minimized the accuracy variance (Fig. S1). Around this optimum (set of SNPs which maximized accuracy and minimized SNP number), prediction accuracy significantly decreased (Fig. S2). The decrease in accuracy induced by using the last SNP sections (least significant SNPs) was accentuated when the number of SNPs was reduced. This confirmed the hypothesis that using the least associated SNPs added noise and spoiled the predictive ability of our genomic prediction model. Indeed, when the number of SNPs increased, the difference in SNP content between the first and last section was reduced. For example, the first and last sections of 2,000 SNPs (on 2,101 SNPs) only differed by 101 SNPs.

In CV2 (incomplete designs) and CV3 (new environments), the training dataset contained all the genetic values, since it included at least one record per variety. The only issue was how to fit the genomic prediction model; there was no new genetic value to predict. We achieved a perfect fit of the model using at least 250 SNPs, whatever the section rank. Since SNPs were first pre-selected based on linkage disequilibrium, the fact that 250 SNPs were sufficient to distinguish 214 varieties appeared logical.

Effect of the number of SNPs and section rank on G × E interactions prediction

In the three cross-validation (CV) designs, G×E interactions (GE_{ij}) were compared to their predictors (gw_{ij}) estimated using only SNPs and environmental covariates (ECs). In all cases, highest accuracies ($r_{(GEij/gwij)}$) were reached using the most significant SNPs (section 1) with a section size of 500 SNPs for NUE and 250 SNPs for NHI

(Fig. 2). Maximum accuracies in CV1, CV2, and CV3 were 0.42 +/- 0.19, 0.53 +/- 0.15, 0.55 +/- 0.30 for NUE and 0.40 +/- 0.20, 0.42 +/- 0.18, 0.38 +/- 0.37 for NHI, respectively. We detected a significant decrease around these optimums in CV1 and CV2 for NUE and especially for NHI (Fig. S3). No accuracy variance patterns were observed.

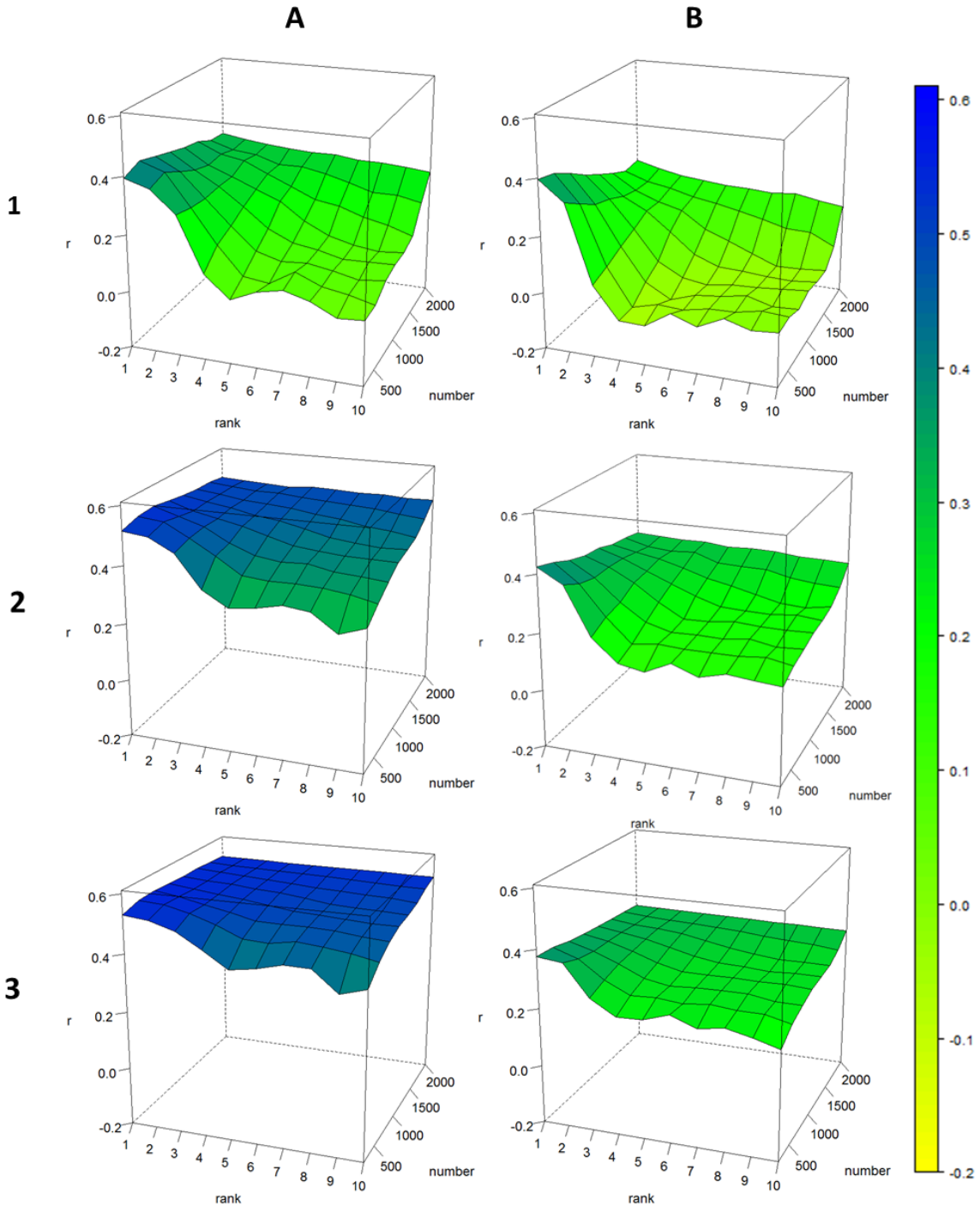


Figure 2: Evolution of G × E interaction prediction accuracy of (A) NUE and (B) NHI in (1) CV1, (2) CV2, and (3) CV3. In CV1 and CV2, predictions were assessed using a three-fold design repeated 50 times. In CV3, a four-fold design was repeated 28 times.

Prediction of environmental values with different sets of SNPs

To predict varieties environmental values, we used two kinship matrices in our G-BLUP-like genomic prediction model: **K1** for genetic values and **K2** for G×E interactions. **K1** and **K2** may share common SNPs. We compared the accuracy of environmental value prediction with and without the G×E predictor (gw_{ij}) (models (6) and (7) in Materials and Methods). We also compared the accuracy between predictions made using all available SNPs (**K1** = **K2**: 25,368 SNPs) and using the optimum.

As previously mentioned, we defined the optimum as the set of SNPs (section size and section rank) which maximized accuracy and minimized SNP number for each matrix. Then, following the previous results, at optimum for NUE, we used 1,275 different SNPs (**K1**:1,250 SNPs; **K2**:500 SNPs) in CV1, and 523 SNPs (**K1**:250 SNPs; **K2**:500 SNPs) in CV2 and CV3. For NHI, we used 700 different SNPs (**K1**:500 SNPs; **K2**:250 SNPs) in CV1, and 322 SNPs (**K1**:250 SNPs; **K2**:250 SNPs) in CV2 and CV3.

Table 2: Effect on accuracy of adding G × E prediction and SNP pre-selection. The number of SNPs used to compute matrices $\mathbf{K_1}$ and $\mathbf{K_2}$ [models (5) and (6); see Materials and Methods] is indicated in columns $\mathbf{K_1}$ and $\mathbf{K_2}$. When all available SNPs were used, $\mathbf{K_1}$=$\mathbf{K_2}$. $r_{(yijk\text{-}Ej/gi)}$ and $r_{(yijk\text{-}Ej/gi+gwij)}$ are the prediction accuracies of models (5) and (6), respectively.

Trait	CV	Optimum				Using all SNPs	
		$\mathbf{K_1}$	$\mathbf{K_2}$	$\mathbf{r_{(yijk\text{-}Ej/gi)}}$	$\mathbf{r_{(yijk\text{-}Ej/gi+gwij)}}$ [a]	$\mathbf{K_1 = K_2}$	$\mathbf{r_{(yijk\text{-}Ej/gi+gwij)}}$ [b]
NUE	1	1,250	500	0.53+/-0.07	0.61+/-0.05***	25,368	0.52+/-0.06***
	2	250	500	0.63+/-0.03	0.72+/-0.02***	25,368	0.71+/-0.02 ns.
	3	250	500	0.61+/-0.07	0.66+/-0.14*	25,368	0.67+/-0.10 ns.
NHI	1	500	250	0.34+/-0.04	0.44+/-0.04***	25,368	0.25+/-0.05***
	2	250	250	0.35+/-0.02	0.46+/-0.03***	25,368	0.41+/-0.03***
	3	250	250	0.31+/-0.06	0.36+/-0.12*	25,368	0.34+/-0.12 ns.

[a] *Result of the Wilcoxon test between $r_{(yijk\text{-}Ej/gi)}$ and $r_{(yijk\text{-}Ej/gi+gwij)}$ at optimum*
[b] *Result of the Wilcoxon test between the optimum and the use of all SNPs for the complete model*
****: P-value < 0.001; **: P-value < 0.01; *: P-value < 0.05; and ns.: non-significant P-value > 0.05*

At optimum, we always achieved a significant improvement in accuracy by adding G × E interaction prediction (Table 2). For NUE, the increase in accuracy ranged from 9 % in CV3 to 15 % in CV1. For NHI, this increase ranged from 16 % in CV3 to 35 % in CV2.

Upon comparing the use of all available SNPs (25,368) and the optimum, we concluded that SNP pre-selection efficiency depended on the cross-validation design and the trait (Table 2). We achieved a significant improvement by pre-selecting SNPs on both traits only in CV1; however, the number of SNPs decreased drastically. Indeed, in CV1, we achieved the highest prediction accuracy of genetic values ($r_{(Gi/gi)}$) only around the optimum (Fig. 1, Fig. S2). In CV2 and CV3, when we used more than 250 SNPs, prediction accuracy of genetic values stayed at the highest level, whatever the section rank and SNP number. Moreover, regarding G × E contribution to genetic

variance (Table 1), the decrease of accuracy around optimum (Fig. 2; Fig. S3) was not sufficient to reveal any significant difference on phenotypic value predictions, except for NHI in CV2.

In conclusion, the first pre-selection that we made based on LD maintained accuracy, although the number of SNPs was reduced (Table S1). Then, depending on CV design, pre-selection based on MET-GWAS results increased accuracy or maintained it, although the number of SNPs was even more reduced.

DISCUSSION

Regarding methodology, previous studies increased the number of SNPs by adding SNPs ranked from first to last. Our study gave a second dimension to the SNP pre-selection issue by independently testing the number of SNPs and the effect of using different kinds of significance. NHI results confirmed the usefulness of this second dimension, given that the optimum did not contain the most highly associated SNPs in our MET-GWAS (Fig. S4). This could also mean that results from our MET-GWAS model may not provide the best overview of genetic architecture.

In GWAS, we control the false positive rate by estimating genotypes' kinship. The goal is to focus only on allelic variation that is linked to the trait, regardless of varieties' genetic background. This can lead to an increase in type II errors (false negatives) if an important part of the genetic variation underlying a trait is linked to genotypes' kinship. Moreover, Rincent et al. [17] recently showed that power of detection can also be reduced in GWAS if the tested SNP is in high LD with various SNPs used to assess the genotypes' kinship. Then, highly significant SNPs result from a balance between SNP effects and their complementarity with the kinship matrix. Moreover, both phenology and end-use quality have a huge impact on several agronomic traits such as NUE [18]. Thus, in GWAS, major genes of phenology and quality are more likely to be linked to agronomic traits and create confounding effects that hide hiding other loci with smaller effect. To deal with this issue, we chose to develop a MET-GWAS model using both phenology and quality information as covariates for NUE. We did not use these covariates for NHI and hypothesized that they may, in part, explain why the optimum section of markers for NHI was not the first one.

The overview of genetic architecture provided by our MET-GWAS results is partially biased. Nevertheless, this study demonstrated the benefit to use them. As an improvement of our method, the improved linear mixed model for GWAS (FaST-LMM-Select [19]), which is able to deal with confounding effects, could be customized to multi-environment data.

The use of genetic architecture information has improved the prediction accuracy of genetic values in human [11], dairy cattle [12], maize [9], and rice [12]. Other studies [7, 8, 10] concluded that excluding least significant markers did not increase prediction accuracy. In these studies, SNP density was perhaps too low to achieve maximum accuracy regarding population size. This is one plausible explanation, although to properly compared studies we would need information on LD in the studied population. Using a training set of 3,305 genotypes (dairy cattle), Vazquez et al. [8] assessed prediction

accuracy for a maximum number of 2,000 SNPs. Hayes et al. [7] tested the same maximum number of SNPs using a training set of 756 genotypes. However, Hayes et al. [20] showed that more than 50 % of adjacent SNPs (studying 38,259 SNPs) had a LD (measured by the usual square of the Pearson correlation) lower than 0.2 in the same panel of Australian Holstein cattle. Zhao et al. [10] tested between 100 and 800 SNPs for a training set of 630 maize genotypes. In comparison, in a study by Schulz-Streeck et al. [9], accuracy decreased when the number of SNPs exceeded 1,750-4,000 (depending on the pre-selection method) with a training set of 2,581 maize genotypes. In the present study, we used up to 2,101 SNPs which were pre-selected based on LD to minimize redundancy, and achieved the highest prediction of genetic values using around 60 % (1,250 / 2,101) for NUE and 24 % (500 / 2,101) of the genome.

Schulz-Streeck et al. [21] modeled G × E interaction with the most consistent SNPs across environments and a relatively small number of markers. These two factors may be the reason why they did not observe much improvement in the prediction of genetic values. In constrat, Heslot et al. [15] selected the most variable SNPs between environments to predict G × E interaction values and achieved an improvement in prediction accuracy. These authors maximized the captured G × E variance using 250 markers. In our study, the best set of SNPs for predicting G × E interaction included 500 markers for NUE, and 250 for NHI. As in Heslot et al. [15], adding more markers to the best set reduced our prediction accuracy.

Genomic prediction methods use a broad diversity of approaches including different assumptions about the distribution of loci effects. This may cause differences between studies. In our model, we assumed a normal distribution of SNPs effects. Some penalized regression approaches such as LASSO mimic pre-selection by leading to sparse solution (some markers had no effects). Thus, it may be reasonable to test our conclusion with a penalized regression approach. However, the number of markers that have an effect depends on the size of phenotypic data set, which can be limiting for complex traits in the context of a multi-environment study.

In this study, prediction accuracy was finally computed with a complete model used to predict environmental values ($r_{(yijk-Ej/gi+gwij)}$). Thus, regarding traits variances decomposition, perfect prediction accuracies would have been 0.84 for NUE and 0.56 for NHI. Indeed, the part of variance explained by the residual error term is not really a genomic prediction issue. This residual variance is influenced by numerous factors such as trial design, soil heterogeneity, model adjustment and accurate measurements. This is mostly a supposedly unpredictable experimental issue that highlights the impact of trial reliability on varieties characterization. However, it can confuse conclusions when studies are compared. In the same way, studies often concluded on the efficiency of their genomic prediction models in accounting for G × E interactions by assessing the gain in accuracy observed when they introduced G × E predictors in their models. However, this gain depends mainly on the part of the variance explained by G × E interaction. Thus, the real issue is not only how to increase prediction accuracy, but

also how to explain G×E interactions as much as possible.

In this study, at optimum, we predicted for NUE and NHI respectively, 17.6 % and 16 % in CV1, 28.1 % and 17.6 % in CV2, and 30.2 % and 14.4 % in CV3 of the G×E variance using 18 ECs. Using 139 genotypes in 340 environments with 68 ECs and 2,395 SNPs, Jarquin et al. [16] reported an increase in accuracy of 17 % and 34 % in CVs implemented in a 10-fold design similar to our CV1 and CV2. G × E variance accounted for 30 % of the total genetic variance (G_i + GE_{ij}). Thus, using published accuracy values, we estimated that around 11.5 % and 31.3 % of the G × E variance were actually predicted. In a CV similar to our CV3 (where we tested the capacity to predict new environments) Heslot et al. [15] used 437 genotypes in 44 environments described with 101 ECs and 250 SNPs and reported an 11.1 % gain in accuracy. G × E variance accounted for 63 % of the total genetic variance. Thus, we estimated that around 8.5 % of the G × E variance was predicted. Nevertheless, the cross-validation design (2-fold) was challenging.

Different genomic prediction approaches can be compared but accuracies and accuracies gain by adding predictor of G × E interactions should always be balanced by the results of the analysis of variance.

This study, which used 214 elite European wheat varieties evaluated in eight environments for NUE and NHI, is the first to demonstrate that SNP pre-selection based on previous knowledge of causal loci can increase prediction accuracy or at least maintain it in a multi-environment framework. Moreover, this study confirmed that G × E interactions can be predicted using molecular information and simple environmental characterization. There are numerous public and private GWAS databases available for different plant and animal species. Their integration in genomic prediction methods is promising for increasing efficiency of genomic selection or personalized medicine.

MATERIALS AND METHODS

Experimental datasets

This study focused on nitrogen use efficiency (NUE) and N harvest index (NHI). The phenotypic data used in this study are described in Cormier et al. [18]. In this study we defined an environment as a combination of year × location × N level. In total, 225 elite European wheat varieties were evaluated in eight environments (two years, three locations, two N levels). In half of the environments, an augmented design was used with four controls. In the other half, all varieties were repeated twice in a complete block design.

In every environment, rainfall, minimum, maximum and average temperature, potential evapotranspiration, and global radiation were measured daily. Eighteen environmental covariates (ECs) were computed based on these measurements (Table S2, Table S3). These ECs are related to nitrogen, drought, heat, and radiation stress throughout the entire plant life cycle or they are focused on specific phenologic stages.

Among the 225 varieties included in field trials, 214 were genotyped using a 90K Illumina chip and SNPs developed by Biogemma. In total, 25,368

SNPs were available in this panel with a minor allele frequency superior above 5 %, no more than 25 % missing data, and no heterozygous loci.

Multi-environment genome-wide association study (MET-GWAS)

The MET-GWAS model was fitted using a mixed model written in R using the ASReml-R package [22]. Following Cormier et al. [18] results, covariates were introduced to avoid quality and precocity confounding effects on NUE; no covariates were introduced for NHI. The model also included an SNP main effect and SNP-by-EC interaction, and was expressed as:

$$y_{ijk} = \mu + e_j + x_i + \alpha_i + \sum_{c=1}^{n} \beta_i ec_c + u_i + \varepsilon_{ijk} \quad (1)$$

$$x_i = \begin{Bmatrix} q_i + bf_i \\ 0 \end{Bmatrix}$$

where y_{ijk} is the phenotypic value of genotype i in the replicate k of environment j, μ is the trait general mean, e_j the effect of environment j, q_i the effect of the quality class of genotype i, b the general sensitivity to flowering time, fi the mean flowering date of genotype i, α_i the allele of genotype i at marker α, β_i the sensitivity of allele α_i to the EC c, ec_c the value of EC c in environment j, $u_i \sim N(0, \sigma_u^2 \mathbf{K})$ a genetic background effect with **K** a matrix of relative kinship, and a residual error term $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$.

K was estimated by a Rogers' Distance [23] matrix based on 3,461 SNPs selected for having less than 0.1 missing data and different genetic map locations.

For each SNP, EC were introduced into the model following a forward approach based on the likelihood ratio test (LRT) using a P-value (P) threshold of 0.05. Then, a Wald test was performed on the complete model to test SNP main effect. LRT and Wald test P-values were used to rank SNPs. We then described how we split these rankings to pre-select the SNPs used in genomic prediction.

Estimation of genetic values and genotype-by-environment interactions

Genomic predictions were first made using a two-step approach to separately assess the effect of SNP pre-selection on genetic value predictions and on G×E predictions.

In the first step, we simply estimated genetic values and G×E interaction values from phenotypic observations. These values were estimated using a model for best linear unbiased estimation and expressed as:

$$y_{ijk} = \mu + E_j + G_i + GE_{ij} + \varepsilon_{ijk} \quad (2)$$

where yijk are phenotypic values, μ the general mean, E_j and G_i are environment j and genotype i fixed effects, respectively, GE_{ji} is the interaction between genotype i and environment j with a residual error term $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$.

Genomic predictions of genetic values and genotype-by-environment interactions

In the second step of the approach, we made genomic predictions of the genetic values and G×E interaction values. We used the model developed by Jarquin et al. [15]: an extension of G-BLUP

implemented in the BGLR package for R [24] which fit reaction norm using reproducing kernel Hilbert space.

Using estimations from equation (2) of genetic values G_i and genotype-by-environment interaction GE_{ij}, we first made independent genomic predictions of G_i and GE_{ij} to clearly identify the optimum set of SNPs (number and ranking in the MET-GWAS) to use in each component.

For genetic value prediction, we computed the following model:

$$G_i = g_i + \varepsilon_{ik} \quad (3)$$

where Gi is the genetic value of genotype i from equation (2), $g_i \sim N(0, \sigma_g^2 \mathbf{K_1})$ with **K1** a genomic relationship matrix and $\varepsilon_{ik} \sim N(0, \sigma_\varepsilon^2)$ a residual error term corresponding to the part of genetic values that is not explained by the marker-based kinship.

When we predicted G × E interactions, we used the model:

$$GE_{ij} = gw_{ij} + \varepsilon_{ijk} \quad (4)$$

where GE_{ij} is the G × E interaction value between genotype i and environment j from equation (2) (Fig. S5), $gw_{ij} \sim N(0, \left[\mathbf{Z_g K_2 Z_g'}\right] \circ \mathbf{\Omega} \sigma_{gw}^2)$ with **Zg** an incidence matrix for the vector of genetic effects, $\mathbf{K_2}$ a genomic relationship matrix, $\mathbf{\Omega}$ an environment covariance matrix based on ECs (Fig. S6, Table S3), and $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ a residual error term corresponding to the part of genotype-by-environment interaction that is not explained.

Genomic predictions of environmental values

Finally, we compared two models that make direct predictions of environmental values to assess the impact on accuracy of adding a G × E predictor (gw_{ij}). To this end, we corrected the observed phenotypic values from the main environment effects, and we computed complete models as:

$$y_{ijk} - E_j = g_i + \varepsilon_{ijk} \quad (5)$$

$$y_{ijk} - E_j = g_i + gw_{ij} + \varepsilon_{ijk} \quad (6)$$

with the previously described terms.

SNP pre-selection

To avoid redundancy in SNP information and reduce computation time, SNP number was reduced based linkage disequilibrium from 25,368 to 2,102 SNPs using the critical LD as a cut-off. Critical LD was assessed following Breseghello and Sorrells [25] and estimated to be $r^2 = 0.23$ in this panel.

In this study we wanted to independently address the effect of SNP number and SNP ranking on MET-GWAS. First, we ranked SNPs by their significance in MET-GWAS. Then, given a fixed number of SNPs, we partitioned this ranking into ten sections with possible overlapping between consecutive sections, from section "rank 1" corresponding to the section of the most significant SNPs to "rank 10" corresponding to the section of the least significant SNPs. To address the SNP number issue, we set SNP section size at 250, 500, 750, 1000, 1250, 1500, 1750, and 2000. In total, we tested 80 combinations of SNP section (10) and SNP number (8).

When we worked on genetic value prediction, we ranked SNPs according to the significance of their main effect and used them in the computation of genomic relationship matrix **K1**. When we worked on G × E predictions, we ranked SNP according to their most significant interaction with ECs and used

them in the computation of genomic relationship matrix **K2**.

Cross-validation design and accuracy

We considered three different cross-validation (CV) designs, each one addressing a different prediction issue. In the first design (CV1), we focused on the ability to predict both additive genetic values and G×E interactions of genotypes that had not been evaluated in any environment. In the second (CV2), we assessed the ability to predict values in an incomplete design. And in the third (CV3), we assessed the ability to predict values in new environments. In CV1 and CV2, we used a three-fold cross-validation design repeated 50 times. In CV1 we randomly chose two-thirds of genotypes present in all environments to train the model and then predict the remaining third. In CV2, we randomly chose two-thirds of the complete data set (214 genotypes × 8 environments) to predict the other third. In CV3, we used a four-fold design, meaning that six environments were used to predict the other two. We tested all 28 environment combinations.

We assessed prediction accuracy as Pearson's product-moment correlation coefficient (r) between the prediction and the genetic values (model (3); $r_{(Gi/gi)}$) or the G×E values (model (4); $r_{(GEij/gwij)}$), on the entire vector. For the three cross-validation designs, the 80 section rank and SNP number combinations were tested. Thus, prediction accuracy can be visualized as a surface plot, in the space defined by section rank and SNP number, using the wireframe function (lattice package in R).

For predicting both genetic values and G × E interactions, we defined the optimum as the set of SNPs (combination of section rank and SNP number) which maximized accuracy and minimized SNP number.

We then assessed the accuracy of model (5) ($r_{(yijk-Ej/gi)}$) and model (6) ($r_{(yijk-Ej/gi+gwij)}$) in two configurations: (1) **K1** and **K2** were computed using the two optimum identified using the two-step approach, and (2) **K1** = **K2** was computed using all 25,368 available SNPs.

AUTHORS CONTRIBUTIONS STATEMENT

Conceived and designed the experiments: FC JLG SP JC. Analyzed and interpreted the data: FC DL JLG SP JC. Wrote the paper: FC DL JLG SP JC.

ACKNOWLEDGMENTS

Data were obtained thanks to the support of the ANR ProtNBle project (06 GPLA016). Analyses were a part of a PhD scholarship (129/2012) supported by the ANRT (Association Nationale de la Recherche et de la Technologie) and a PhD scholarship supported by the INRA metaprogramme SelGen and by the Auvergne region. Data were obtained thanks to the work of the research staff at the Estrées-Mons experimental unit (INRA) and at Villiers-le-Bâcle and Vraux (Arvalis). The authors are also grateful to the Biometrics and Statistics Unit of CIMMYT (Mexico) for hosting this work. Sincere thanks to Paulino Perez for his help on the BGLR package, to Brigitte Mangin for her advice on interpretation and comments on the manuscript, and to Alma McNab for English proofing.

REFERENCES

1. Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Sci 48:391-407.
2. Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29.
3. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.

4. Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative trait in maize. Crop Science 47:1082-1090.
5. Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166-177.
6. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. Genetics 185:1021-1031.
7. Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME (2010) Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-Fat Percentage, and Type in Holstein Cattle as Contrasting Model Traits. PLoS Genet 6(9): e1001139. doi:10.1371/journal.pgen.1001139.
8. Vazquez AI, Rosa GJM, Weigel KA, De los Campos G, Gianola D et al. (2010) Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins. J Dairy Sci 93:5942-5949.
9. Schulz-Streeck T, Ogutu JO, Piepho HP (2011) Pre-selection of markers for genomic selection. BMC Proceedings doi:10.1186/1753-6561-5-S3-S12.
10. Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP et al. (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769–776.
11. de los Campos G, Vazquez AI, Fernando R, Klimentidis YC, Sorensen D (2013) Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. PLoS Genet 9(7): e1003608. doi:10.1371/journal.pgen.1003608.
12. Zhang Z, Ober U, Erbe M, Gao N, He J, et al. (2014) Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association Studies. PLoS ONE 9(3): e93017. doi:10.1371/journal.pone.0093017.
13. Cooper M, DeLacy IH (1994) Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theor Appl 88:561-572.
14. Burgueño J, De los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci 52:707-719.
15. Heslot N, Akdemir D, Sorrells ME, Jannink JL (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor Appl Genet 127:463-480.
16. Jarquin D, Crossa J, Lacaze X, Du Cheyron Philippe, Daucourt J et al. (2014) A reaction norm model for genomic selection using high‐dimensional genomic and environmental data. Theor Appl Genet 127:595-607.
17. Rincent R, Moreau L, Monod H, Kuhn E, Melchinger AE et al. (2014) Recovering power in association mapping panels with variable levels of linkage disequilibrium. Genetics 197:375-387.
18. Cormier F, Faure S, Dubreuil P, Heumez E, Beauchêne K et al. (2013) A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (Triticum aestivum L.). Theor Appl Genet 126:3035–3048.
19. Listgarten J, Lippert C, Kadie CM, Davidson R, Eskin E, et al. (2014) Improved linear mixed models for genome-wide association studies. Nat Methods 9: 525–526.
20. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433–443.
21. Schultz-Streeck T, Ogutu JO, Gordillo A, Karaman Z, Knaak C et al. (2013) Genomic selection allowing for marker-by-environment interaction. Plant Breeding 132:532- 538.
22. Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) ASReml-R reference manual. Queensland Department of Primary Industries. http://www.vsni.co.uk/downloads/asreml/release2/doc/asreml-R.pdf
23. Roger JS (1972) Measures of genetic similarity and genetic distance. In: Studies in Genetics VII. Austin: University of Texas publication. pp. 145-153.
24. de los Campos G, Perez-Rodriguez P (2013) BGLR: Bayesian generalized linear regression. R package version. http://R-Forge.R-project.org/projects/bglr/
25. Breseghello F and Sorrells M (2006) Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172:1165-1177.

AND EPISTASIS IN ALL OF THAT?

In the previous paper, we wanted to extend classical GWAS and G-BLUP models by taking into account SNP × environmental covariates (EC) interactions. The goal was to make models come closer to biology. However, in order to be as close to biology as possible, there is something that is rarely modelized: epistasis.

Epistatic interactions have been recognized to be a fundamental component of the understanding of (i) the structure and function of genetic pathways, (ii) the evolutionary dynamics of complex genetic systems and (iii) the genetic variance (Cheverud and Routman 1995; Carlborg and Haley 2004; Mackay et al. 2014). If large scale epistasis analyses become much more systematic in yeast or animal species, these approaches are still under prospected in plants. The main limitation of comprehensive analyses is the total number of interactions that must be studied. Nevertheless, the development of new methods (Cordell 2009) and afforadable informatic hardware (*e.g.* calculator) make it possible to launch pioneer studies in plant too.

Here we will discuss the preliminary work made on epistatic interactions aiming (once finished) to (i) identify epistatic genes, (ii) dissect epistatic networks and (iii) integrate these results in genomic prediction. We addressed epistatic interactions from a statistical point of view at the level of population (non-additivity of loci effects).

Is it important?

The first question we should answer is: "How much of genetic variance epistatic interactions explain?" Indeed, if it is a really small proportion in nitrogen use efficiency (NUE) related traits, it may not be worth considering the computing challenge.

Due to its size and its composition, our dataset does not allow us to assess this proportion. Nevertheless, GWAS results of part IV (Cormier et al. 2014) may help to address this issue. In the "predictions" section of the previous paper, we assessed the adjusted mean variance explained by summing quantitative trait nucleotide (QTN) effects (r^2_{adj}, Table 4, part III). We can compare it to the sum of individual QTN prediction accuracy (sum of r^2_{snp}) (Fig. 3). This difference can be impressive. For example, for plant height the sum of r^2_{snp} and r^2_{adj} were equal to 177.7 and 48.6 %, respectively; in agreement with a high epistatic contribution in the genetic control of plant height assessed by several authors and already mentioned (see discussion in the paper part III). At the opposite, for straw dry

matter (ADM_S) the sum of r^2_{snp} and r^2_{adj} were equal to 110.8 and 52.8 %, respectively, using the same number of QTN as for plant height.

This difference between the sum of r^2_{snp} and r^2_{adj} resulted from missing data, addition of misestimation of QTN effects, redundancy between information [*i.e.* linkage disequilibrium (LD) between QTN and number of QTN] and epistatic interactions between QTN. We did not assess any significant difference between traits in missing data neither in mean LD between QTN. Thus, we computed the ratio [r^2_{adj} / sum of r^2_{snp}] only corrected for the number of QTN and first hypothesized that it will be mainly related to epistatic interactions. In agreement with this hypothesis, we assessed a negative correlation ($P < 0.01$, $r^2 = 0.19$) between this corrected ratio and trait mean r^2_{snp}. Indeed, when epistatic interactions are high the power of detection decreases (as narrow-sense heritability decreases) leading to the detection of only bigger QTN, and thus to an increase in the mean r^2_{snp}. But, there is a second plausible explanation. The proportion of shared information between QTN (r^2 = LD) did not vary between traits, but the proportion of genetic variance explained and shared by QTN increased with QTN effects. Thus, a trait controlled by large effect loci, will have a higher mean r^2_{snp} and a smaller (r^2_{adj} / sum of r^2_{snp}) corrected ratio.

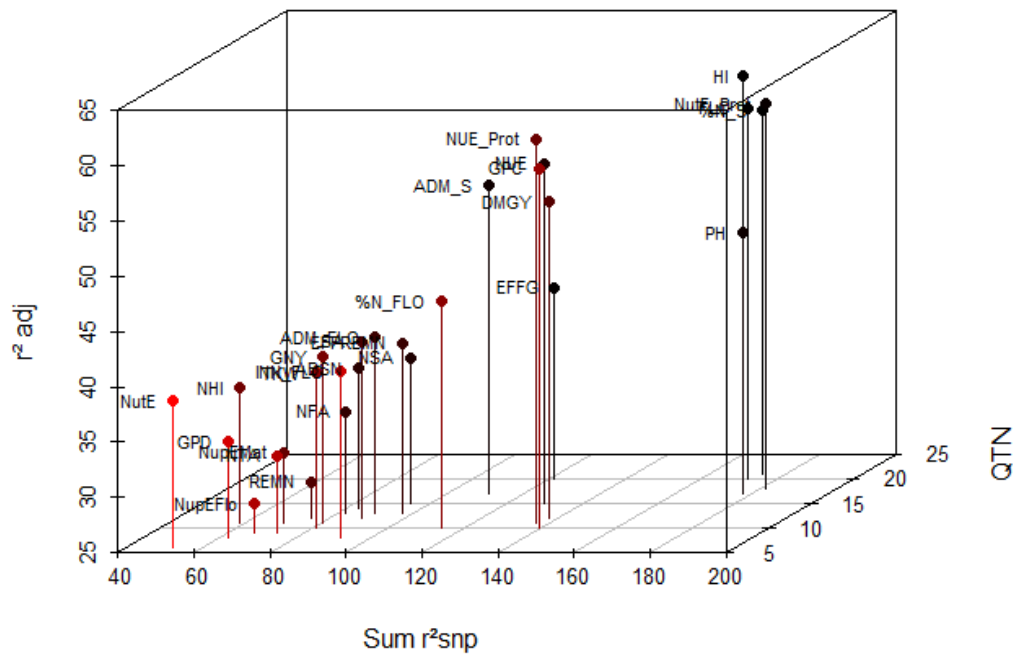


Figure 3: 3D plot of r^2_{adj} in function of Sum of r^2_{snp} and QTN number for 28 traits.

In our dataset, having an *a priori* on the contribution of epistasis in traits variance may not be possible. Moreover, this discussion is limited to epistatic interaction between QTN leaving aside interactions between SNP that do not have a significant additive effect. Nevertheless, several studies revealed that epistasis cannot be ignored when describing the genetic architecture of complex traits (for a review

Mackay 2014). Thus, given the number of genotypes in our dataset, we should first focus on two-way epistatic interactions (SNP × SNP) and *a posteriori* quantify their contributions.

Genome-wide detection of epistatic interaction

For additive loci and loci interacting with environmental covariates, we made a whole-genome detection and we demonstrated that this knowledge could be useful in both deciphering pathways and increasing the efficiency of genomic prediction models. For epistatic loci, we kept the same methodology and focused our work on nitrogen use efficiency (NUE; grain yield / NTA_{max}). Thus, we started by a step of whole-genome detection using an extension of the classical genome-wide association study (GWAS) model K (Fig. 4). The goal was to find networks of epistatic interactions significantly involved in NUE in our panel (Fig. 4) and to identify the involved metabolic pathways.

We decided to use the model species *A. thaliana* to build our networks as information on pathways is reduced in wheat, while the database of protein-protein interactions, transcription factors and co-expressions are much more developed on *A. thaliana*.

First, we blasted all our markers context sequences or anchors to the *A. thaliana* genome and conserved only SNP located in putative wheat paralogs. Then, we tested the significance of SNP pairwise interactions. And finally, we compared these interactions to the ones registered in *A. thaliana* interactome databases using the paralogs genes on which SNP may be located (Fig. 4).

This allowed (i) to reduce the number of tested interactions to the ones that we were able to screen in *A. thaliana* interactome database. Five hundred days of computing (10 days on 50 CPU) were already necessary to achieve the pairwise detection. (ii) It decreased the confounding effect of LD between SNP. In fact, highly interconnected sub-networks tend to be group of SNP in high linkage disequilibrium (*e.g.* left of Fig. 4). (iii) At the end, it allowed to draw a simplified network based on gene (instead of SNP) containing less false positive interactions. Indeed, significant interactions from our extended GWAS model K performed on wheat NUE are cross-validated by experimental or computing approaches on completely unrelated data. Once again, we chose not to be too stringent on significance threshold [-log(P-value) > 3] and to cross-validate using various sources of information. However, we should keep in mind that wheat and *A. thaliana* are phylogenetically distant. Thus, common interactions may be reduced to conserved pathways among plant species. More generally, we now have a dataset of significant SNP interaction that can be used in MAS models.

We ended this work with a small interaction network (right of Fig. 4) that required further investigations. Indeed, this network is composed of "validated" interactions (*e.g.* Suppl.data 10) explaining a significant part of NUE variance in our panel (r^2 = 6.5 +/- 3.84 % of the genetic variance). Added to that, we also may have identified the genetic markers linked to the causal polymorphism involved in the interaction. A branch of this network is particularly interesting as it contains the Ferredoxin-Dependent Glutamine-Oxoglutarate Aminotransferase (Fd-GOGAT) gene and several genes involved in photorespiration, nitrogen assimilation and senescence.

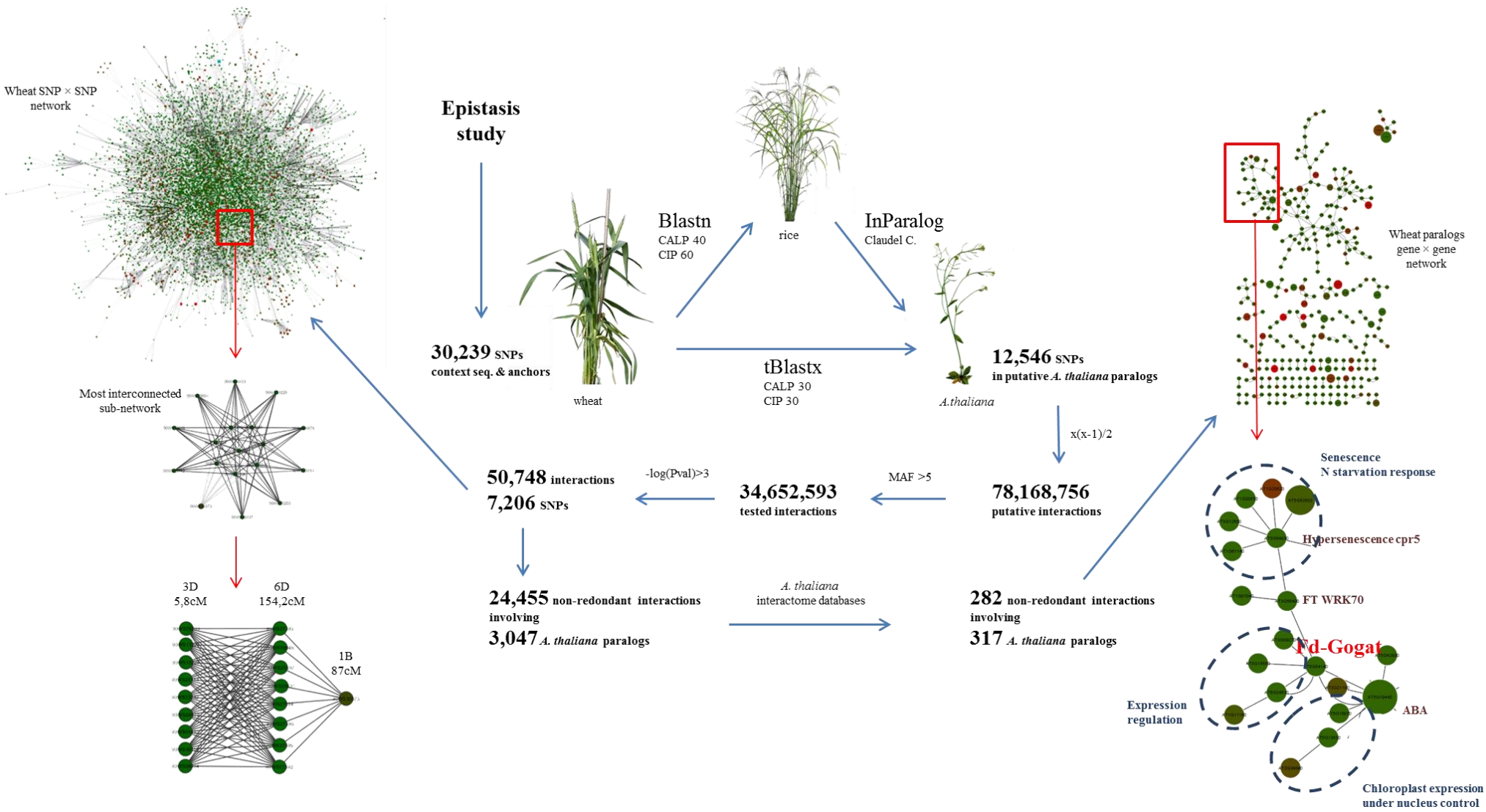


Figure 4: Diagram of epistatic interactions analyses. The model used to test epistatic interaction was the following: $y_{ijk} = \mu + e_j + \alpha_i + \alpha_i' + \alpha_i \alpha_i' + u_i + \varepsilon_{ij}$ where y_{ijk} is the phenotypic value of genotype i environment j (dataset described in Cormier et al. 2014), μ is the trait general mean, ej the effect of j, α_i and α_i' the alleles of genotype i at marker α and α', $ui \sim N(0, \sigma_u^2 \mathbf{K})$ a genetic background effect with **K** a matrix of relative kinship, and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ a residual error term. *A. thaliana* interactome databases were requested through CORNET using the co-expression (Pearson correlation coefficient > 0.8), protein-protein interaction (experimental and predicted) and transcription factor (confirmed and knock-out experiments) modules.

Adding epistatic interaction in GS model

In G-BLUP, we hypothesize that the distribution of markers' effects follows a normal distribution with a homogeneous variance between them. But, individual effects of markers are not directly estimated and used in prediction. Predictions are based on genotypes kinship assessed from genomic information (Meuwissen et al. 2001). In our GS model, kinship matrices ($\mathbf{K_1}$, $\mathbf{K_2}$ and $\mathbf{K}$) were mainly related to the probability of having a common allele as we computed kinship matrices following the formula:

$$\mathbf{K_{inship}} = \frac{[\mathbf{MatInc} \times \mathbf{MatInc}']}{n_{SNP}} \qquad (7)$$

with n_{SNP} the total number of SNP used to compute $\mathbf{K_{inship}}$ (= $\mathbf{K_1}$, $\mathbf{K_2}$ or $\mathbf{K}$) and **MatInc** a genotyping matrix converted to a centered and reduced incidence matrix (number of genotypes × n_{SNP}).

Therefore, there is also a part of epistasis in the information contained in our kinship matrices. Indeed, the number of common epistatic interaction (**N**) between two varieties can be described as a function of the probability to have a common allele between two genotypes (approximated by **K**):

$$\mathbf{N} \approx \sum_{i=2}^{n} \binom{n}{k} \mathbf{K_{inship}}^{i} \quad (8)$$

with k the order of epistatic interaction (number of involved loci) and n the total number of SNP.

However, this part of epistasis, which is already taken into account in our model, rapidly becomes null. Indeed, for two genotypes, the probability of having the same epistatic interaction is the probability of having a common allele to the power of the interaction order ($\mathbf{K_{inship}} < 1$; $\lim_{i \to n} \mathbf{K_{inship}}^{i} = 0$). We also advocate building kinship matrices not based on an overview of the genome, but on SNP having additive effects and SNP having additive effects interacting with environmental covariates (EC), which may reduce even more the part of epistasis as we focused on a subset of the total genotyping data.

More generally, Gianola et al. (2006) suggested that non-parametric GS models (*e.g.* reproducing kernel Hilbert spaces) compared to parametric models (*e.g.* G-BLUP) would be better suited to take into account epistatic contribution in trait genetic architecture. Comparing 10 parametric models to four non-parametric models, Howard et al. (2014) confirmed that non-parametric models over performed when genetic architecture was based entirely on epistasis. However, this simulation study was only based on two-way epistatic interactions and parametric models were slightly better for additive genetic architecture. Added to that, this kind of study compared basic models which were not especially customized to integrate epistasis.

In multi-environmental dataset, SNP additive effects and SNP × environmental covariates (EC) interactions need to be estimated. If we add epistatic interactions, the number of estimations increases

even more. Thus, we may retain our G-BLUP approach which is not limited by the number of estimated effects.

To integrate epistasis, we could extend the approach that we had on genotypes-by-environment (G × E) interactions. Indeed, when we integrated prediction of G × E interactions we added the predictor gw_{ij} defined as:

$$gw_{ij} \sim \mathrm{N}(0, [\mathbf{Z_g K_2 Z_g'}] \circ \mathbf{\Omega} \sigma^2_{\mathrm{gw}}) \quad (9)$$

with $\mathbf{Z_g}$ an incidence matrix for the vector of genetic effects, $\mathbf{K_2}$ a genomic relationship matrix, $\mathbf{\Omega}$ an environment covariance matrix based on EC.

We could add a similar predictor named gg_i:

$$gg_i \sim \mathrm{N}(0, [\mathbf{Z_g K_e Z_g'}] \circ [\mathbf{Z_g K_e Z_g'}] \sigma^2_{\mathrm{g}}) \quad (10)$$

with $\mathbf{K_e}$ a genomic relationship matrix based on SNP involved in epistatic interactions.

But pairwise interactions will not be conserved. In fact, with a Hadamard product, we will modelize all the interactions between all SNP. This was also the case of our predictor of G × E interactions (gw_{ij}). The fact that a SNP could interact with a particular EC was not conserved and we modelized the response of SNP to all EC.

In fact, if we want to conserve the information on pairwise interactions, we may have to work directly on the way we assess kinship between genotypes and base this kinship on the selected interactions.

We could define a modified kinship ($\mathbf{K_{modif}}$) as:

$$\mathbf{K_{modif}} = \frac{1}{\mathrm{n_{inter}}} \sum_{\mathrm{i=1}}^{\mathrm{n_{inter}}} [\mathbf{MatInc_{i1}} \times \mathbf{MatInc_{i1}}'] \circ [\mathbf{MatInc_{i2}} \times \mathbf{MatInc_{i2}}'] \quad (11)$$

with $\mathrm{n_{inter}}$ the number of SNP × SNP interactions, $\mathbf{MatInc_{i1}}$ and $\mathbf{MatInc_{i2}}$ the genotyping matrices (of i1 and i2 the two interactors of interaction i, respectively) converted to incidence matrices (number of genotypes × 1).

And finally, our predictor would become:

$$gg_i \sim \mathrm{N}(0, [\mathbf{Z_g K_{modif} Z_g'}] \sigma^2_{\mathrm{g}}) \quad (12)$$

with the terms previously described.

Preliminary results showed accuracy improvement (Table 3) for predictions of new genotypes when the epistasis predictor (model 10) was added to the complete model of predictions (model (6) in the previous paper). However, when this predictor was computed using all SNP [model (10) based on **K** instead of $\mathbf{K_e}$] accuracy did not increase. Thus, here again, SNP pre-selection may be useful.

Table 3: Effect on accuracy of adding epistasis predictor selecting SNP × SNP interactions at a significance threshold = 5. Significance of SNP × SNP interactions were calculated using the model described in legend of Fig. 4.

	Model	Matrices	Content	Accuracy[a]
Optimum base	**(6)**	**K1**	1,250 SNPs	0.62+/-0.04
		K2	500 SNPs	
+ SNP×SNP predictor	**(10)**	**K_e**	832 SNPs	0.65+/-0.04 ***
		K	25,368 SNPs	0.62+/-0.04 ns.
	(12)	**K_{modif}**	1380 Interactions	0.62+/-0.04ns.

[a] *Result of the Wilcoxon test between the accuracy at optimum and the accuracy when a SNP × SNP predictor is added*
****: P-value < 0.001; **: P-value < 0.01; *: P-value < 0.05; and ns: non-significant P-value > 0.05*

Computing interactions between all pairs of interactive SNP [Haddamart product; model (10)] appeared more effective than keeping pairwise interaction information (**K_{modif}**) at a –log(P-value) threshold = 5 . When we selected SNP to build SNP and SNP × EC predictors, we saw that adding even SNP with low significance in MET-GWAS increased accuracies (Fig. S4). We can hypothesize that it may be the same for epistasis interaction and need to test really less stringent significance thresholds.

In any case, more investigations are required on both genome-wide mapping of epistatic loci and epistasis integration in GS. Nevertheless, we have a really interesting dataset to start investigation on pathways and to support the development of new methods. One of the challenges in this type of work is to deal with huge dataset that cannot be processed using R. Thus, we should also pursue a development of methods using the C++ coding language.

References

BreedWheat project: http://www.breedwheat.fr/index.php. Work package 2: Genetics and ecophysiology of wheat adaptation to biotic and abiotic stress in the frame of sustainable agricultural systems and global climate change http://www.breedwheat.fr/fr/page.php?optim=WP2

Carlborg O, Haley C (2004) Epistasis: too often neglected in complex trait studies? Nature Rev Genetics 5:618-625.

Cheverud JM, Routman EJ (1995) Epistasis and its contribution to genetic variance components. Genetics 139:1455-1461

Cordel HJ (2009) Detecting gene-gene interactions that underlie human disease. Nature Rev Genetics 10:392-404

Cormier F, Le Gouis J, Dubreuil P, Lafarge S, Praud S (2014) A genome-wide identification of chromosomal regions determining nitrogen use efficiency components in wheat (Triticum aestivum L.). Theor Appl Genet, 127: 2679-2693.

Gianola, D, Fernando RL, Stella A (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics, 173:1761-1776.

Mackay TFC (2014) Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nature Reviews Genetics, 15:22-33.

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157:1819-1829.

Howard R, Carriquiry AL, Beavis W (2014) Parametric and Nonparametric Statistical Methods for Genomic Selection of Traits with Additive and Epistatic Genetic Architectures. G3, 4:1027-1046.

Yu JM, Pressoir G, Briggs WH, Vroh Bi, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208.

GENERAL DISCUSSION

APPLICATIONS IN BREEDING

Impact of past selection

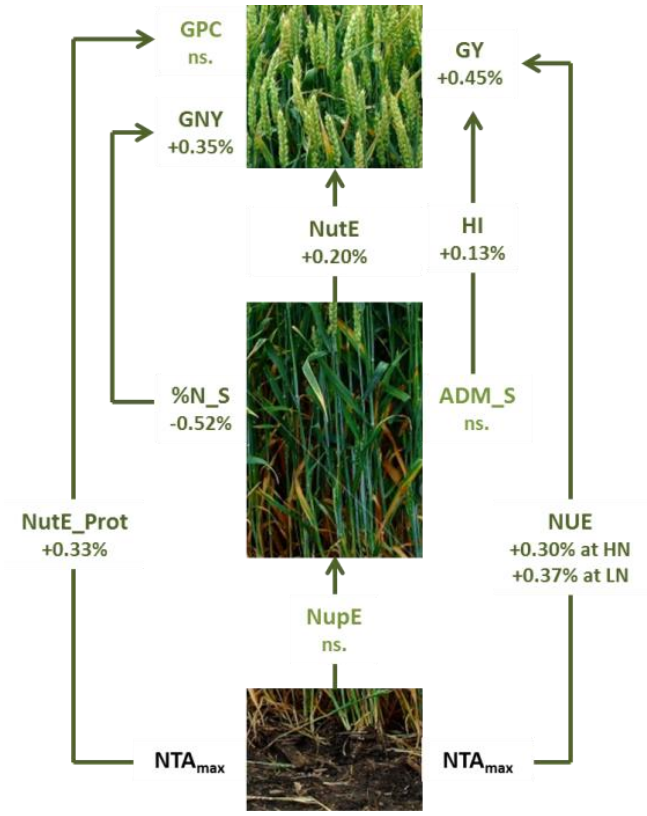


Figure 1: Summary of past genetic progresses assessed in Part II.

We have shown that nitrogen (N) use efficiency (NUE) genetic improvement was driven by direct selection on grain yield (GY) while maintaining grain protein content (GPC) quite stable. There was no consensus on the components of GY increased by selection (grain weight, number of grains per spike or number of spikes per area). We were not able to test changes in N uptake (NupE) due to the absence of a sufficient genetic variance for this trait with the variance decomposition model of Part II. Thus, NUE was increased by a better N partitioning meaning that N utilisation efficiency (NutE) was improved. During the selection process, genotypes were tested in numerous experiments where moderate N stresses surely occurred leading to improved GY stability hence NUE at low N regimes.

The impact of this past selection was also highlighted during the genome-wide association study (GWAS). Indeed, focusing on varieties released from 2005, we assessed a significant correlation between frequencies of alleles having a positive effect on a trait and the genetic correlation between this trait and GY (Discussion and Suppl. data 7, Part III). This led to a high median allele frequency of favourable additive alleles for traits under this GY-driven selection.

To conclude on past breeding, three challenges appear: (i) to accelerate the genetic progress by combining favourable additive loci regarding their putative epistatic and environmental interactions, (ii) to increase

uptake efficiency and (iii) increase protein concentration in low N environments. We provided variances decompositions in multi-environmental trials, genetic markers tagging chromosomal regions with additive effects and/or interacting with environmental covariates (EC) and/or with other chromosomal regions and models for multi-environmental genomic prediction. These are useful tools to face future challenges. Nevertheless, these tools should be validated on another dataset such as the dataset generated by the on-going BreedWheat project, in which 103 varieties are common with our dataset and where similar N regimes have been in different site × year combinations. This will be completed in a few months.

Phenotypic selection

In our dataset, we detected significant genotype × environment (G × E) and genotype × N regimes (G × N) interactions leading to an indirect selection efficiency of 78.1 % for NUE regardless of the selection type (*i.e.* in high N for low N or in low N for high N; Part II). This confirms previously published works (Breeding for NUE, Part I) and leads us to conclude that indirect selection does not overpass direct selection. On this basis, we recommend selecting at the targeted N regimes. In a context of fertiliser reduction, this targeted N regime is the low N (*i.e.* in fact, suboptimal in our study). To develop new wheat varieties is a long task. Thus, breeders should ideally start to select in low N environment as soon as possible. Having said that, indirect selection efficiency for NUE is high. And selecting at high N, they are already selecting for low N environment with a relatively good efficiency. However, the reciprocal argument can also be used to advocate for selection in low N environment, even if varieties will be cultivated at high N.

Few arguments give sense not to start selection at low N. One of them is that bread-making quality could be mis-estimated at low N as protein concentration is significantly decreased. This problem mainly arises from the fact that only few breeding programs are selecting for a higher bread-making quality and/or grain protein concentration and the majority are selecting for grain yield while maintaining quality.

Increasing grain protein concentration and bread-making quality at low N is a major constraint on an effective N reduction. A combined approach using genetics and agronomy may help to solve this issue. For example, we can hypothesize that delaying the last N fertiliser supply (Bogard et al. 2010) or the introgression of *NAM-A1a* in elite germplasm could be a part of the solution.

An affordable compromise between double trial (high N and low N) and single trial (high N) could be to preselect a reduced number of genotypes and to test them also in low N conditions. This type of segmentation is already made to assign genotypes to northern or southern France trials regarding their precocity. As phenotyping for NUE is more complex, we may use genotypic information. We can imagine building a prediction model based on effects detected in genome-wide association studies (GWAS; multi-environmental and epistatic) to preselect these genotypes or to use genomic selection (GS) methods. Both approaches have to be properly compared. Indeed, our results do not permit to directly compare their efficiencies. Using GWAS results we explained on average 29.7 % ($r = 0.54$) of NUE within environment

variance with 15 additive SNP. With our GS model using only additive predictor (g_i), for new genotypes (cross-validation 1), we had a prediction efficiency of r = 0.53. These results can appear similar, but explicative and predictive efficiencies cannot be compared. Here again further investigation on an independent dataset should be performed. Biogemma has all the dataset to do so.

Until now, the first thing a breeder should implement is an efficient way to assess the level of N available in their trials. In this way, they could at least put their yield measurements in regards to N stresses and could start to classify genotypes by N regimes. As previously mentioned in Part II, using control varieties for which total N will be assessed could be a cost effective solution. However, a main limitation will be the confounding effects of others environmental stresses (*e.g.* heat and drought stresses).

Changing NUE genetic architecture

Phenotypic selection focused on the final conversion of N into grain and did not enhance all NUE components. Major improvement has been made on NutE. Consequently, past and new varieties have significantly different NutE genetic values leading to a high NutE genetic variance. In our panel, NutE was heritable and powerful QTL detection could be performed providing genetic markers facilitating the combination of favourable alleles. In contrast, NupE heritability was low meaning that phenotypic selection cannot be performed efficiently. Moreover, for traits with a relative low heritability, we showed that GWAS results are rapidly becoming useless to predict environmental values. Finally, breeding for an enhanced NutE is easier as it has already been improved. And to enhance NupE seems hardly possible as it has been neglected. We seem stuck in the past breeding framework. Something needs to be changed in NUE genetic architecture. It could be done by adapting (i) the way we select or (ii) the germplasm that we used.

Concerning the way we select, the challenge is to better balance selection pressure among NUE-related traits. We need to counterbalance the impact of the GY-driven selection. Marker-assisted selection (MAS) on traits hardly phenotyped and/or with weak heritability may be useful. Pre-anthesis N status (INN_FLO) is a good example. Indeed, in Part III, we showed that INN_FLO QTL had major pleiotropic effects on NUE-related traits. Thus, we concluded that we should focus on this trait. However, regarding phenotyping difficulties and its intermediate heritability of 0.63, we can understand that INN_FLO is not used in breeding program. Nevertheless, INN_FLO genetic variance exits. And using the methodology of part II to assess past genetic progress, it appears that INN_FLO has never been improved. Among the seven QTL for INN_FLO discovered in part III, three were not associated with flowering date. Consequently, MAS for enhanced INN_FLO without affecting regional adaptation is possible and should be tested.

For low heritability traits such as NupE, it is more complex. Indeed, low heritability could result from an actual low genetic variance compared to the total phenotypic variance or from weak measurement accuracy. Anyway, on this panel, phenotypic selection is nearly impossible with our phenotyping method. In contrast, marker-assisted selection can be efficient. In fact, phenotypic and marker-assisted selections act

at different scales. Variance on causal genes may exist but the way these causal genes may be combined in varieties can result in similar breeding values. First, we should focus on the genetic variance itself and dissect its genetic determinism. Then, we will try to combine QTL to increase the part of genetic variance on the total phenotypic variance. In this sense, NupE (= NupEMat in part III) additive QTL have been detected and tools to combine them in elite lines are available. We assessed a low median allele frequency of favourable alleles at these QTL (0.33 in varieties registered after 2005). And the few related varieties (Andino, Uski, Premio, Isengrain) with a significant number of NupE favourable alleles had an enhanced NUE. Thus, improvement is here again possible and necessary.

Regarding the germplasm used in selection, the issue is to know if alleles with a major impact on NUE could be found in a wider diversity. Introduction of dwarfing alleles (*Rht* genes) is a good example of alleles from an exotic diversity answering to global agronomic issues (*e.g.* Peng et al. 1999). Indeed, in the context of the Green Revolution, demand for high-yielding varieties able to deal with an increased fertiliser application was achieved through their introduction. Coming from a broader diversity, *Pch1*, an eyespot (*P. herpotrichoides*) resistance gene from *Aegilops ventricosa* (Mena et al. 1992) commonly used in American and European breeding programs, is also a good example. The work-package three of the BreedWheat project completely fits in this approach as one of its tasks is to explore a broad genetic diversity to bring new favourable alleles in elite germplasms. We could also imagine finding causal genes using GWAS-based QTL cloning or transcriptomic analyses. Then, we could screen different germplasms (*e.g.* exotic, mutants, related species) to look for new alleles of these specific genes that could enhance NUE-related traits once introgressed in elite germplasm. Transgenesis can also be a way of creating a new diversity with major effect adapted to elite germplasm by introgressing genes or alleles that are not present in the wheat genome and/or changing regulation of wheat genes.

To conclude, the impact of the GY-driven selection can be counterbalanced using MAS based on alleles coming from elite or more exotic germplasms and affecting neglected NUE-related traits. One of the main questions is also to know which traits can be simultaneously increased. Indeed, for example, even if no antagonist additive QTL were detected between NutE and NupE; it does not mean that no antagonist mechanism exits at all as varieties genetic values for these traits were negatively correlated in our dataset (r = -0.32).

In any case, if MAS has to be performed, we first need to identify causal genes. Thus, questions about methods used in gene discovery arise. Our work already provided new insights. Nevertheless, these methods can be improved starting with our statistical approaches.

IMPROVING METHODOLOGY

Statistical models

The main limitation of our statistical approach is that we did not take into account non-independence of factors very well. Indeed, the use of variance-covariance (VCOV) matrices in our statistical models was not optimized. In GWAS and MET-GWAS, the same kinship matrix was used whatever the trait studied and the SNP tested. Moreover, in genome-wide epistasis detection, we used the same kinship matrix but did not add any specific VCOV matrix for SNP × SNP effects. More generally, we did not use any VCOV matrices for environments (year × site × N regimes) effects in our multi-environmental models (Part II and IV), neither for effects of G × E interactions.

Regarding VCOV matrix for varieties' additive effects, recent studied showed that kinship matrices may have to be computed regarding causal loci for the studied trait and eliminating SNP in LD with the tested one (Listgarten et al. 2014; Rincent et al. 2014; Wang et al. 2014a). Thus, we should improve our GWAS models in this sense. In the same way, the use of the $\mathbf{K}_{\mathbf{modif}}$ described in Part IV could be tested to modelise VCOV for epistatic effects as it is computed regarding only significantly interacting pairs of SNP. We mostly use the ASReml-R package v3.0.1 (Butler et al. 2009) in which several models of VCOV are already available (Boer et al. 2007) and can fit more or less hardly. Thus, regarding VCOV matrix for environmental effects, these models of VCOV should be tested. We could also imagine directly setting VCOV values by computing a VCOV matrix based on environmental covariates. This is actually the kind of matrix that we used in the genomic selection models (matrix $\mathbf{\Omega}$). Thus, in agreement, $[\mathbf{Z_g K Z_g}']^{\circ}\mathbf{\Omega}$ could be used to modelize the VCOV of the G × E interactions.

Although for environments, this is much more complex than for varieties. Indeed, to assess varieties kinship, we used SNP detected in a broad genetic diversity (90K; Wang et al. 2014b) and SNP developed by Biogemma. Added to that, SNP were selected for non-redundancy. Thus, even if 30% of SNP (Biogemma SNP) were detected in a more reduced diversity, the bias induced by SNP in the computation of kinship matrix may be reduced. In contrast, to quantify environmental stresses, we consciously chose the environmental covariates we were interested in. Moreover, the reduced number of environments created significant correlations between environmental covariates. Added to that, all environmental covariates were used. Consequently, our $\mathbf{\Omega}$ matrix was biased by the choice we made to focus on some stresses (bias of selection) and the number of covariates that we calculated for each stress (bias of redundancy/weight). In the same way, Jarquin et al. (2013) did not select environmental covariates to compute $\mathbf{\Omega}$, contrary to Heslot et al. (2014) who performed a first step of environmental covariates pre-selection. Nevertheless, we used environmental covariates based on wheat physiological knowledge covering a wide range of stresses. Added to that, this matrix $\mathbf{\Omega}$ allowed to increase prediction accuracy. Thus, even biased, we can reasonably hypothesize that $\mathbf{\Omega}$ contains useful information that could be used to set environments VCOV values or at least VCOV starting values.

The use of optimized VCOV matrices will improve, but also complicate, our statistical approaches. In our case, the dataset was fixed and computation time was not an issue. Yet in breeding program, dataset are much bigger and computation time impacts reactivity above all when genotypes have to be selected in a short period [between harvest (mid-July) and sowing (mid-October)]. Our MET-GWAS model took 150 hours to test the effects of around 25K SNP on 1 trait. The GWAS model for epistasis detection took 10 days with 50 central processing units to test the effects of around 34K interactions on 1 trait. Thus, speeding up analyses may be a useful improvement. Up to now, FaST-LMM-Select (Listgarten et al. 2014) appears promising as (i) it solves computational issues (dataset size and computational time); (ii) it can be used for epistasis detection; and (iii) it adapts varieties relationship to the trait and chromosomal region studied. However, it is not suited for multi-environmental analyses and the way SNP are selected to derive a rank-reduced relationship between varieties need to be improved (Wang et al. 2014a).
More generally, even an improved model has its limitations. A key point may be to understand them to be able to combine different statistical approaches and different sources of knowledge.

Gene discovery strategy

Usually, once we have selected an interesting QTL, we densify the chromosomal region using SNP mostly developed in genic regions. Then, we declare that the best candidate gene is the one carrying the most significant SNP in GWAS. Consequently, regarding the simulation study (Part II and III) in which causal SNP were randomly chosen among SNP that did not participate to the panel structuration, this approach may be correct in 2/3 of cases. But is this efficiency enough and close to the reality?
Indeed, precise SNP densification can require intensive bioinformatics and lab work in a non-sequenced species such as wheat (*e.g.* reconstruction of the genomic sequence of the region, SNP detection). Moreover, candidate genes may be validated using a genetically modified (GM) approach or used as selection tools by breeders. Thus, we cannot be satisfied with a method having an efficiency of 2/3 on such a decisive step. Moreover, in our simulation study, we may use two false hypotheses: (i) causal mutation did not participate to panel structure and (ii) allelic frequencies in our genotyping dataset were representative to allelic frequencies of causal mutations.
As previously discussed, our GWAS models were not perfect and the way we computed kinship matrix influenced SNP significance making results highly dependent of SNP allelic distribution (*i.e.* frequency and repartition among varieties). Thus, if a causal SNP has unbalanced allele frequencies and/or allelic distribution among varieties related to the panel structure; we can expect that other SNP (having a more homogeneous distribution and a sufficient LD with the causal one) will be more significant in GWAS. In agreement to this, preliminary results showed that a causal SNP linked to the panel structure is not the most significant SNP in its chromosomal region in 75% of cases. The most significant had a mean LD of $r^2 = 0.7$ with this causal SNP. This situation may be frequent among causal mutations determining our studied traits. Indeed, we worked on an historical elite panel of varieties selected for different environment.

Consequently, alleles can be specific to a regional adaptation (linked to the panel structure) and/or newly introgressed in elite germplasm and/or being eliminated (unbalanced frequency). Allelic distribution of causal genes determining traits under selection pressure may be more frequently unbalanced (frequency and distribution among varieties) than expected.

Therefore, further investigations are required, but for traits known to be under selection pressure (*e.g.* GY, NutE), causal mutations may not be randomly distributed along the genome and may be more likely located in chromosomal region under selection pressure and/or involved in the panel structuration. Thus, the proportion of causal mutations being under significance peak may be even less than the estimated and insufficient rate of 2/3. Significance in GWAS should be taken into account but should not be the only criterion to choose candidate genes. This choice has to be better thought and should take into account results of several GWAS (*e.g.* additive, interacting with EC, epistatic), linkage disequilibrium, allelic distribution and previous knowledge on genes located in the associated regions.

We may also adapt our choice of candidate genes to its future utilization. Indeed, genes used in MAS and in GM approach may be different. In fact, in MAS, the goal is to apply an identified effect on a new germplasm. And in GM approach, the goal is to create a new effect in an identified germplasm. Thus, our results may be used differently: while the most significant QTN/QTL (additive, interacting with EC and other SNP) may be the one that will be used in MAS including GS; they may not be the best choices for GM approach. Indeed, effects that we detected directly depend on the phenotypic and genetic diversity. Although in GM approach, the goal is to create a new diversity. Nevertheless, knowing that a gene has already an effect on a trait, we can hypothesize that changing its expression/regulation will have an effect too. However, the detected and the created effects may be unrelated. GM approach often target hub in metabolic pathway. In this sense, epistasis network may be a source of information complementary to GWAS results. In agreement to this, analyses of SNP network based on epistatic interactions revealed that SNP connectivity (number of epistatic interactions) was negatively correlated to the significance of SNP additive effect [$x = 14.5 - 0.62 \times -\log(\text{P-value})$; $P < 0.05$] and/or to the significance of the SNP × environmental covariates interactions [$x = 14.5 - 0.68 \times -\log(\text{P-value})$); $P < 0.001$]. Thus, hubs in SNP epistasis networks have a central role in the traits we studied. However, these hubs would not have been found out if epistasis interactions had not been studied (*i.e.* in GWAS no effects were revealed for SNP tagging these hubs).

Regarding the difficulties linked to the choice of candidate genes, an improvement of our strategy could be to reduce QTL size to focus on fewer genes. However, QTL mean size (3.2cM) was already smaller than expected from simulation (7.8cM). Increasing the threshold used to declare a SNP-trait association significant (QTN) is not a good idea, we showed that increasing the -log(P-value) threshold of QTN significance decreases QTL size. But, it also drastically reduced the power of detection, resulting in a higher proportion of false positive QTL among all computed QTL. Thus, we need to succeed in decreasing QTL size without affecting power of detection. Maybe we can focus on the position of the most significant SNP by chromosomal region and then, from this position, we can compute QTL using the local LD decay

(Fig. 2). In fact, with the method that we used (Fig. 8, Part III), if we had a long distance LD, the first boundaries delimited a long QTL. Then, boundaries were well extended as we took into account this long distance LD for a second time. We may over correct for LD. Nevertheless, we should keep in mind that at least for 1/3 of positive QTL, the most significant SNP was not the one closest to the causal mutation. Thus, we really need to first test the efficiency of this method, after fixing issues linked to our simulation study hypothesis.

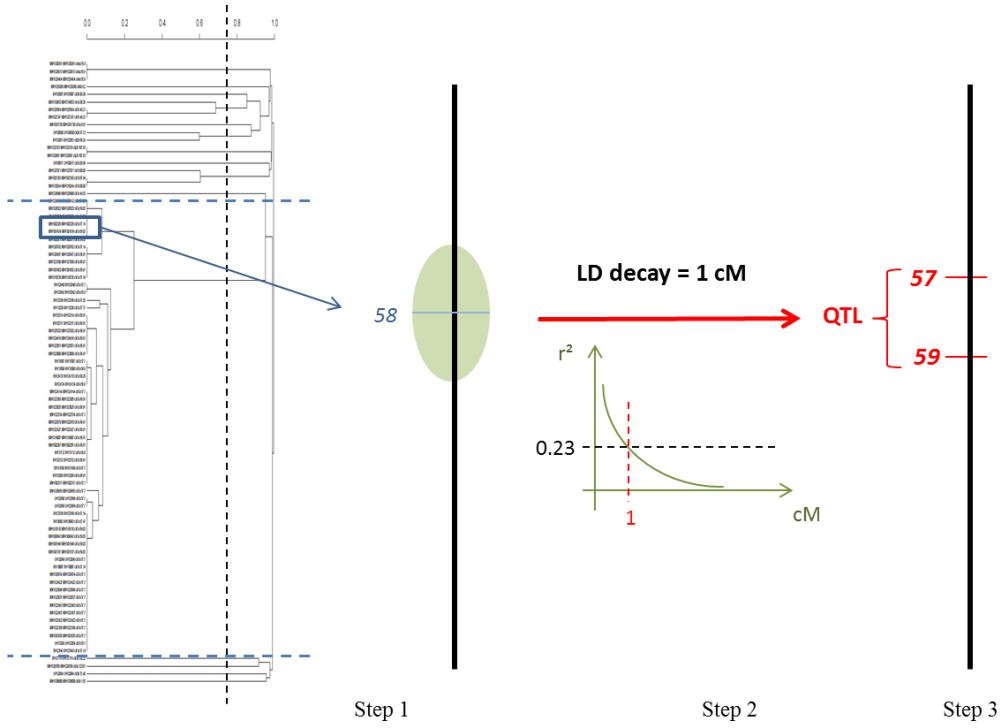


Figure 2: Method that should be tested to define QTL from GWAS result. Step 1: QTN clustering in function of LD (r^2) (method average, cut-off = 1- critical LD). Step 2: Estimation of LD decay around the most significant QTN. Step 3: Creation of QTL boundaries.

Regarding hypothesis made in our QTL definition method, improvement can also be done. Indeed, it is mainly based on one parameter: the critical LD that we used to cluster quantitative trait nucleotide (QTN) and to assess local LD decay. This parameter was set at the 95th percentile of the unlinked r^2 (assessed between two SNP mapped on different chromosomes). However, selection along with other factors can create linkage disequilibrium (LD) between chromosomal regions located on different chromosomes. Thus, our estimation of the critical LD is biased. Due to selection, our panel is also not at the drift-recombination equilibrium required for the function used to assess LD decay [*i.e.* curvilinear function proposed by Hill and Weir (1998)]. Moreover, for this function, the effective population size was set at the panel size. Although, varieties were not totally independent (kinship).

To conclude on the gene discovery strategy, our work provided new insights and tools to diagnose strategy weaknesses. However, improvements can be achieved. Gene discovery strategy needs to be thought in light of limitations of GWAS approaches. And choice of candidate genes should be done compiling GWAS

results, linkage disequilibrium, allelic distribution, previous knowledge on genes located in the associated regions and genes end-uses. However, if we have to combine so many information: which ones should we used? How should we prioritize them? What is redundant and what is not?

Complementarity, redundancy, and choice

Combining different criteria to make a choice can be tricky. Indeed, the number of situation to deal with rapidly increases with the number of information and their complexities. For example, if we want to base the choice of candidate genes on the following four criteria: additive effects, epistatic and environmental interactions, and previous knowledge, with only two simple modalities (significant or not) by criterion; we will already have to deal with 16 scenarios. And for each scenario, we will have to decide what are the further investigations required or the end-use of the candidate genes.

Concerning our results, we already made some choices on some candidate genes and interesting chromosomal regions. However, we need to develop a less subjective approach. The idea is to list all the criteria used to identify the different scenarios. And then, we will properly determine the future of results fitting in each scenario, taking into account that all information may not carry the same weight in the decision.

This also leads to the need to quantify the part of redundancy and complementarity between information. This dilemma can be illustrated by several examples in our work. Colocalisation between our QTL and published functional candidate genes can reveal that these candidate genes are also good candidate in our germplasm (complementarity). However, in our genotyping dataset, SNP are not homogeneously distributed among the genome and chromosomal regions containing published candidate genes contain more SNP. However, these regions were purposely densified in SNP by Biogemma. Thus, there is a higher probability to identify QTL in these previously published chromosomal regions (redundancy). In our GS models, we use SNP tested in a GWAS performed on the same dataset (redundancy). Thus, our results need to be validated in another dataset (complementarity). However, if the genetic, phenotypic and environmental diversities are completely different we may never succeed. We tested random overlaps of information when we tested if QTN colocalisation between traits were significant or not. In the same way, we need to develop methods to quantify or test complementarity/redundancy. Descriptive and analytic statistic can be used. But, here again, we will have to make some choices.

We validated some of them by a step of risk assessment such as the threshold used to declare that a SNP-trait association is significant. Or, we made some of them to simplify the analysis through the use of approximation or assumptions such as the use of the function proposed by Hill and Weir (1998) to assess the LD decay. Some of them were even less consciously made, such as the use of statistical approaches based on the restricted maximum likelihood (REML). In any case, researchers have always made and will still make choices. The main issue is to know the different options and their respective consequences.

GENERAL CONCLUSION

Table 2: Summary of methods and results.

Analyses	Traits	Methods	Results
Genetic progress	17 traits	Use of precocity, quality and height as covariates Decomposition of the genetic progress (G and G × N)	NUE was improved at both HN and LN regimes through selection yield leading to an increase N partitioning
GWAS	28 traits	Method to define QTL from QTN Analyses of colocalisation	333 QTL with additive effect Selection affected QTN distribution Pleiotropic effect of INN_FLO QTL
MET-GWAS	NUE	Use of precocity and quality as covariates Test of SNP × EC interactions	1,240 QTN with additive effect 1,122 QTN interacting with EC
NAM-A1		Allelic distribution and 3D conformation	In elite germplasm, the introgression of the functional allele of *NAM-A1* may improve N remobilisation
Epistatic GWAS	NUE	Whole-genome detection combine with interactome database	7,206 SNP involved in 50,748 epistatic interactions A « validated » gene network of epistatic interactions involved in wheat NUE
GS	NUE NHI	Use of MET-GWAS in GS Effect of SNP pre-selection independantly of SNP number	Using a G-BLUP approach, SNP pre-selection increases prediction accuracies in multi-environments trials

Past breeding effort improved NUE in wheat at both high N and low N regimes. Regarding future challenges, LN seems to be the new targeted regimes. However, varieties were mostly selected regarding yield and all NUE components were not improved in the same way. Thus, breeding method should be adapted to maintain the past breeding effort and re-balance selection pressure among traits. To achieve this purpose, the use of phenotypic selection combined with genotypic selection based on our results may be useful. With this work, we provided tools to facilitate the transition from a breeding in high N to low N and accelerate genetic progresses (Table 2). However, these tools need to be validated in another dataset and investigations in a wider genetic diversity must not be neglected (Table 3).

During this PhD, new methods and new insights in gene discovery strategies were also developed. These methods and strategies can still be improved (Table 3) keeping in mind that changes should be tested to properly assess their impact along the entire pipeline of analyses: from QTN detection to candidate gene identification. The main conclusion of our methodological work is that several sources of information should be used to choose candidate genes. QTN significance should not be the only one and a lot of information has to be cross-referenced. Now, the main issue is to clearly determine how these data should be combined.

Table 3: Summary of improvements and further investigations

In ?	What ?	How ?
Breeding	Select in low N	Increase GPC
	Assess N availability in breeders trials	Assess NTAmax on control varieties
	Balance selection pressure among NUE-related traits	Mix phenotypic and marker-assisted selections
	Study a wider genetic diversity	BreedWheat WP2
Statistical methods	Validate GWAS results	Use of the BreedWheat dataset
	Compare QTL and GS approaches	
	Integrate VCOV matrix in statistical models	Test of $\mathbf{K}_{\mathbf{modif}}$, $\mathbf{\Omega}$ and $[\mathbf{Z_g K Z_g}']^{\circ}\mathbf{\Omega}$
	Speed up GWAS analyses	Review of package and software available
Gene discovery	Decrease QTL size	Test new QTL definition methods
	Rationalize candidate gene choice	List criteria and establish a decision tree

REFERENCES

Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007) A mixed-model quantitative trait (QTL) analysis for multi-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in Maize. Genetics, 177:1801-1813.

Bogard M, Allard V, Brancourt-Hulmel M, Heumez E, Machet JM, Jeuffroy MH, Gate P, Martre P, Le Gouis J (2010) Deviation from the grain protein concentration-grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. J Exp Bot 61:4303-4312

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) ASReml-R Reference Manual. Queensland Department of Primary Industries. http://www.vsni.co.uk/downloads/asreml/release2/doc/asreml-R.pdf

Heslot N, Akdemir D, Sorrells ME, Jannink JL (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor Appl Genet 127:463-480.

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54-78

Jarquin D, Crossa J, Lacaze X, Du Cheyron Philippe, Daucourt J et al. (2014) A reaction norm model for genomic selection using high‐dimensional genomic and environmental data. Theor Appl Genet 127:595-607.

Mena M, Doussinault G, Lopez-Brana I, Aguaded S, Garcia-Olmedo F, Delibes A (1992) Eyespot resistance gene Pch-1. H-93 wheat lines. Evidence of linkage to markers of chromosome group 7 and resolution from the endopeptidase locus Ep-D1b. Theor Appl Genet, 83:1044-1047.

Listgarten J, Lippert C, Kadie CM, Davidson R, Eskin E, et al. (2014) Improved linear mixed models for genome-wide association studies. Nat Methods 9: 525–526.

Peng JR, Richards DE, Hartley NM, Murphy GP, Devos KM, et al. (1999) "Green revolution" genes encode mutant gibberellin response modulators. Nature 400: 256–261.

Rincent R, Moreau L, Monod H, Kuhn E, Melchinger AE, et al. (2014). Recovering power in association mapping panels with variable levels of linkage disequilibrium. Genet 197:375-387

Wang Q, Tian F, Buckler ES, Zhang Z (2014a) A SUPER powerful method for genome wide association study. PloS One Sep 23;9(9):e107684.

Wang S, Debbie W, Forrest K, Allen A, et al. (2014b) Characterization of polyploidy wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotech J doi: 10.1111/pbi.12183

ANNEXES

SUPPORTING INFORMATION ON PART II

[Supplemetary data of Cormier et al. (2013) (2013) A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat *(Triticum eastivum* L.). Theor Appl Genet 126:3035–3048]

Supplementary data 1: Trials descriptions: locations (A), treatments (B) and fertilisations (C)

A)

	VB08	VR09	EM08	EM09
Year	2008	2009	2008	2009
City	Villiers-le- Bacle	Vraux	Estrées-Mons	
Latitude (°N)	48.72	49.02	49.88	
Longitude (°E)	2.17	4.23	3.04	
Elevation (m)	200	85	85	
Previous crop	Rapeseed		Oats	
Sowing date	17/10/2007	23/10/2008	22/10/2007	23/10/2008
Harvest date	24/07/2008	06/08/2009	30/07/2008	03/08/2009
Plot size (m²)	10		6.5	
Sowing density (grains m^{-2})	250	320	240	
Clay (%)	25	45	16.6	15.2
Loam (%)	70	55	75.5	72.4
Sand (%)	5	0	7.9	11.7
Rain during crop cycle (mm)	487	525	493	390

B)

	VB08	VR09	EM08	EM09
Treatment type 1	Anti-Slug	Herbicide	Anti-Slug	Herbicide
Treatment name 1	TDS premium	first	Extralugex 5R	Defi
Treatment dose 1	3.0 kg ha^{-1}	0.8 l ha^{-1}	4 kg ha^{-1}	4.25 l ha^{-1}
Treatment date 1	31/10/2007	19/11/2008	23/10/2007	24/10/2008
Treatment type 2	Herbicide	Herbicide	Herbicide	Herbicide
Treatment name 2	Quartz GT	Atlantis	Defi	Gratil + Allié
Treatment dose 2	2.2 l ha^{-1}	300 g ha^{-1}	5 l ha^{-1}	36 g ha^{-1} + 30 g ha^{-1}
Treatment date 2	22/01/2008	14/04/2009	25/10/2008	02/04/2009
Treatment type 3	Fongicide	Growth regulator	Growth regulator	Growth regulator
Treatment name 3	Opus	Stabilan	Mondium	Mondium
Treatment dose 3	0.5 l ha^{-1}	2 l ha^{-1}	2.5 l ha^{-1}	2.5 l ha^{-1}
Treatment date 3	18/03/2008	10/04/2009	31/03/2008	14/04/2008
Treatment type 4	Fongicide	Growth regulator	Herbicide	Fungicide
Treatment name 4	Unix	Moddus	Gratil + Allié	Unix + Opus
Treatment dose 4	0.8 kg ha^{-1}	0.2 l ha^{-1}	22 g ha^{-1} + 22 g ha^{-1}	0.8 kg ha^{-1} + 0.5 l ha^{-1}
Treatment date 4	18/03/2008	24/04/2009	03/04/2008	22/04/2009
Treatment type 5	Herbicide	Fongicide	Fungicide	Fungicide
Treatment name 5	Atlantis WG	Flexity	Unix + Opus	Joao + Twist 500sc
Treatment dose 5	0.26 kg ha^{-1}	0.4 l ha^{-1}	0.8 kg ha^{-1} + 0.6 l ha^{-1}	0.8 l ha^{-1}+ 0.2 l ha^{-1}
Treatment date 5	04/04/2008	24/04/2009	04/04/2008	19/05/2009
Treatment type 6	Growth regulator	Fongicide	Fungicide	Fungicide
Treatment name 6	Moddus	Opus	Joao + Twist 500sc	Caramba
Treatment dose 6	0.5 l ha^{-1}	0.5 l ha^{-1}	0.8 l ha^{-1} + 0.2 l ha^{-1}	1.1 l ha^{-1}
Treatment date 6	25/04/2008	24/04/2009	05/05/2009	12/06/2009
Treatment type 7	Fongicide	Fongicide	Fungicide	Insecticide
Treatment name 7	Virtuose+Joao	Bravo 500	Caramba	Karaté K
Treatment dose 7	0.4 l ha^{-1}	1 l ha^{-1}	1.4 l ha^{-1}	1 l ha^{-1}
Treatment date 7	25/04/2008	13/05/2009	30/05/2008	29/06/2009
Treatment type 8	Fongicide	Fongicide	Insecticide	
Treatment name 8	Amistar	Menara	Karaté K	
Treatment dose 8	0.5 l ha^{-1}	0.5 l ha^{-1}	1 l ha^{-1}	
Treatment date 8	30/05/2008	13/05/2009	12/06/2008	
Treatment type 9	Fongicide	Fongicide		
Treatment name 9	Caramba Star GC	Epopée		
Treatment dose 9	0.5 l ha^{-1}	0.9 l ha^{-1}		
Treatment date 9	30/05/2008	03/06/2009		
Treatment type 10		Insecticide		
Treatment name 10		Karaté Zéon		
Treatment dose 10		0.08 l ha^{-1}		
Treatment date 10		28/05/2009		

C)

	VB08		VR09		EM08		EM09	
	HN	**LN**	**HN**	**LN**	**HN**	**LN**	**HN**	**LN**
Soil residual N date	11/02/2008		20/02/2009		15/02/2007		07/02/2008	
Soil residual N (kg N ha-1)	106	106	30	30	67	67	30	30
N fertilisation date 1	28/03/2008		24/02/2009		06/03/2008		16/03/2009	
N fertilisation rate 1 (kg N ha-1)	66.5	44	60	60	50		50	0
N fertilisation date 2	23/04/2008		26/03/2009		26/03/2008		21/04/2009	
N fertilisation rate 2 (kg N ha-1)	60	0	100	60	70	70	50	50
N fertilisation date 3			18/05/2009		28/04/2008		30/04/2009	
N fertilisation rate 3 (kg N ha-1)			60	0,00	50	0	50	0
Estimation of %N_S	DUMAS		DUMAS		NIRS		NIRS	
Estimation of GPC	NIRS		NIRS		NIRS		NIRS	

Supplementary data 2: Year of release (YR), quality, mean height (PH) and precocity (FLO) of wheat varieties tested. Quality grade are the common breadmaking classes used by the National Association of French Millers: BAF, very high quality; BPS, high quality, BP, good quality, BA, biscuit quality, and BAU, other use. PH and FLO are varieties genetic BLUEs. Precocity is characterized by the day of flowering (GS65, anthesis half way) after the 1st January. PH are in cm. YR comes from the French) and the European catalogue of agriculture species.

Variety	YR	Quality	FLO	PH	Variety	YR	Quality	FLO	PH	Variety	YR	Quality	FLO	PH
ACCOR	2007	BPS	132	81	CRAKLIN	1998	BB	136	80	ORQUAL	1991	BPS	139	74
ACIENDA	2004	BPS	135	73	CROUSTY	1995	BB	140	94	ORVANTIS	2000	BPS	140	80
ACIENTO	2007	BPS	138	76	DIALOG	2008	BP	141	79	PACTOLE	1987	BPS	139	88
ACONEL	2007		143	85	DINOSOR	2005	BPS	141	75	PAINDOR	1996	BPS	145	73
ADEQUAT	2006	BPS	148	80	DSV_50115			148	76	PAJERO	1995	BP	142	101
ADONIS	2007		147	80	DUXFORD	2006	BPS	147	76	PALADAIN	2006	BPS	142	77
AGRESTIS	2002	BP	146	80	EINSTEIN	2002	BPS	144	74	PALEDOR	2005	BB	136	82
AGUILA	2005	BP	136	73	EM07162			139	85	PARADOR	2000	BPS	146	82
ALCAZAR	2004	BP	145	77	EMERALD	2007	BPS	145	75	PAROLI	2004	BPS	146	87
ALDRIC	2007	BPS	137	85	ENESCO	1996	BPS	132	72	PEPIDOR	2007	BP	143	89
ALEZAN	2007	BPS	138	74	EPHOROS	2004	BP	144	98	PERFECTOR	2004	BPS	145	79
ALFA	2008		149	86	EPIDOC	2006	BPS	135	77	PERICLES	2005	BAU	143	78
ALIGATOR	2010	BPS	136	77	EQUILIBRE	2003	BPS	139	83	PHARE	2008	BPS	143	74
ALIXAN	2005	BPS	137	77	ESPERIA	2002	BAF	132	81	PIKO	1994		150	89
ALLISTER	2003	BP	140	79	ESTICA	1991	BAU	148	87	POTENZIAL	2006	BPS	146	84
ALTIGO	2007	BP	138	82	ETECHO	1994	BP	134	80	PR22R20	2002	BPS	146	75
ALTRIA	1996	BAU	135	83	EUCLIDE	2007	BPS	136	81	PR22R28	2000	BP	143	78
AMBITION	2005	BAU	149	83	EVEIL	2003	BPS	137	73	PR22R58	2002	BPS	134	73
AMERIGO	2002	BPS	138	82	EXELCIOR	2007	BPS	136	80	PREMIO	2007	BPS	138	77
AMUNDSEN	2008	BP	148	77	EXOTIC	2005	BP	135	78	QUALITY	2002	BAF	134	67
ANDALOU	2002	BP	135	77	EXPERT	2007	BP	144	81	QUATUOR	2002	BPS	137	66
ANDINO	2007	BPS	135	78	FARANDOLE	1999	BP	139	78	RAISON	2006	BP	147	78
ANTILLE	2006		136	81	FIORENZO	2002		133	67	RASPAIL	2002	BPS	147	80
ANTONIUS	2006	BAF	144	101	FIORETTO	2008	BPS	136	83	RECITAL	1986	BPS	133	79
APACHE	1998	BPS	137	77	FLAIR	1996	BAU	147	93	RENAN	1989	BAF	140	88
ARACK	2006	BPS	141	75	FORBAN	2002	BP	145	81	RESSOR	2004	BB	137	76
ARCHE	1989	BAU	139	80	FRELON	2001	BP	139	81	RICHEPAIN	2006	BPS	140	73
AREZZO	2008	BPS	136	81	GALACTIC	2007	BAU	137	71	RITMO	2004	BAU	148	81
ARLEQUIN	2007	BPS	137	80	GALIBIER	1992	BAF	133	90	ROBIGUS	2002	BAU	147	76
ASTRAKAN	2003	BPS	139	78	GARANTUS	2007	BP	147	83	RODRIGO	2006	BPS	134	73
ASTUCE	2004	BPS	146	84	GARCIA	2006	BP	134	78	ROSARIO	2004	BP	147	78
ATTLASS	2004	BP	142	87	GLASGOW	2003	BB	145	72	ROYSSAC	2002	BPS	135	77
AUBUSSON	2002	BPS	136	76	GRAINDOR	2006	BPS	135	87	RUBENS	1995	BP	140	89
AUDI	2005		148	83	GRETHEL	2008	BP	136	76	RUNAL	1998	BAF	142	85
AURELE	2003	BPS	147	78	GUADALUPE	1997	BPS	133	80	RUSTIC	2005	BP	137	77
AUTAN	2001	BPS	134	70	GUARNI	2004		134	79	SAMURAI	2005	BAU	147	80
AUTENTIC	2007	BPS	145	76	GULLIVER	2005	BPS	147	77	SANKARA	2004	BPS	142	77
AVANTAGE	2005	BP	145	93	HARDI	1969	BPS	143	93	SATURNUS	2001	BAF	143	96
AXIMACK	2007	BPS	146	80	HATTRICK	2001	BP	146	83	SCIPION	1982	BP	137	77
AZIMUT	2004	BPS	136	77	HAUSSMANN	2006	BPS	146	82	SEBASTO	2007		141	75
AZTEC	1994	BPS	136	76	HYPERION	2005		149	72	SELEKT	2007	BPS	144	85
AZZURO	2006	BPS	141	83	INCISIF	2005	BPS	145	81	SEYRAC	2006	BPS	147	80
BAGOU	2007	BB	139	76	INOUI	2004	BP	136	71	SHANGO	1994	BPS	147	85
BASTIDE	2003	BPS	136	78	INSPIRATION	2006	BP	147	87	SIGNAL			144	95
BATTANT	2006	BAU	146	85	INSTINCT	2006	BPS	138	77	SIRTAKI	2007	BPS	135	74
BERMUDE	2007	BPS	141	83	INTERET	2008	BPS	144	89	SISLEY	1998	BP	139	77
BISCAY	2000	BAU	147	77	IRIDIUM	2007	BPS	142	82	SOCCER			145	85
BOISSEAU	2007	BP	143	78	ISENGRAIN	1997	BPS	137	78	SOGOOD	2006	BPS	145	80
BOKARO	2003		134	77	ISIDOR	2002	BP	134	77	SOISSONS	1988	BPS	135	79
BOLOGNA	2002	BAF	134	76	ISTABRAQ	2003	BAU	146	82	SOLLARIO	2007	BPS	135	78
BOREGAR	2008	BPS	139	78	JB_ASANO	2008	BPS	144	88	SOLUTION	2007	BP	143	79
BOSTON	2001	BAU	144	78	KALANGO	2002	BPS	134	73	SOPHYTRA	2007	BP	146	87
BOTTICELLI	2004		134	81	KORELI	2007	BPS	142	86	SPECTRO	2007		144	85
BUENNO	2008	BP	135	82	LANCELOT	2002	BPS	147	78	SPONSOR	1995	BP	144	91
CABELLO	2007		141	85	LEU_88-02-1			144	84	TALDOR	1997	BPS	135	80
CALISTO	2002	BPS	139	78	LIMES	2002	BP	146	87	TAMARO	1997	BAF	145	81
CAMP_REMY	1980	BPS	141	87	LONA	1997	BAF	137	92	TAPIDOR	2002	BAU	138	83
CAMPARI	2003	BAU	148	79	MANAGER	2006	BP	148	92	TEXEL	1992	BP	139	83
CAMPERO	2006	BPS	138	81	MARKSMAN	2006		143	75	TIAGO	2008	BPS	138	82
CAPHORN	2001	BPS	140	74	MAXWELL	2007	BAU	141	77	TIFOSO	2008		136	70

CAPNOR	2001	BP	144	81	MELKIOR	2004	BPS	143	81	TIMBER	2005	BP	143	78
CAPO	1997	BAF	144	110	MENDEL	2004	BPS	138	78	TOGANO	2004	BAF	144	89
CARIBOU	2006	BPS	143	77	MENESTREL	2007	BPS	137	79	TOISONDOR	2004	BP	142	70
CARNAVAL			136	73	MERCATO	2005	BPS	137	75	TOREADOR	2002	BPS	145	83
CCB_INGENIO	2006	BPS	133	81	MESSAGER	1994	BAU	138	86	TREMIE	1992	BAU	136	81
CEZANNE	1998	BPS	136	85	MESSIDOR	2007	BP	137	77	TROCADERO	2002	BP	132	84
CHAGALL	2004	BP	144	78	MH_05-32			138	80	USKI	2009	BAU	137	80
CHARGER	1997	BPS	142	76	MINOTOR	2007	BPS	141	75	VALODOR	2007	BPS	135	81
CHEVALIER	2006	BPS	146	84	NIRVANA	2001	BPS	140	74	VANTORIS	2007		138	75
CIGALO	2007		137	70	NUAGE	2006	BPS	142	78	VERLAINE	2007	BPS	144	78
CLAIRE	1997	BAU	147	79	OAKLEY	2006	BAU	146	73	VISCOUNT	2007	BAU	147	72
CM2713			145	80	OCTET	2007	BPS	136	76	VM9601			146	85
COMODOR	2008	BPS	142	83	OEDIPE	2007	BP	141	83	WALDORF	2006		147	80
COPERNICO	2004		133	73	ORATORIO	1996	BP	138	80					
CORDIALE	2005	BPS	141	72	ORNICAR	1997	BB	140	79					
CORVUS	2000	BP	146	88	ORPIC	1998	BPS	136	82					

Supplementary data 3: Phenotypic traits descriptions

Trait	Description	Formula	Units
FLO	anthesis date		days (after 1st January)
PH	plant height		cm
ADM_S	straw dry matter at maturity		kg ha^{-1}
%N_S	straw N content at maturity		%
SA	spike per area		nb spike m^{-2}
TKW	1000-kernel weigth		g
GY	dry matter grain yield		kg ha^{-1}
GPC	grain protein concentration		%
NSA	straw N per area	ADM_S × %N_S	kg ha^{-1}
KS	kernel per spike	GY / (TKW × SA)	nb kernel per spike
GNY	grain N yield	GPC / 5,7 × GY	kg ha^{-1}
NTA	total N per area	NSA + GNY	kg ha^{-1}
HI	harvest index	GY / (GY+ADM_S)	%
NHI	N harvest index	GNY / NTA	%
NupE	uptake efficiency at maturity	NTA / NTA_{max}	%
NutE	utilisation efficiency	GY / NTA	kg DM kg^{-1} N
NUE	N use efficiency	GY / NTA_{max}	kg DM kg^{-1} N
GPD	grain protein deviation	GPC - a × GY -b (a and b are trial properties)	% of protein
NutE_Prot	N utilisation efficiency to protein	GPC / NTA	% protein kg^{-1} N ha^{-1}
NUE_Prot	N use efficiency to protein	GPC / NTA_{max}	% protein kg^{-1} N ha^{-1}

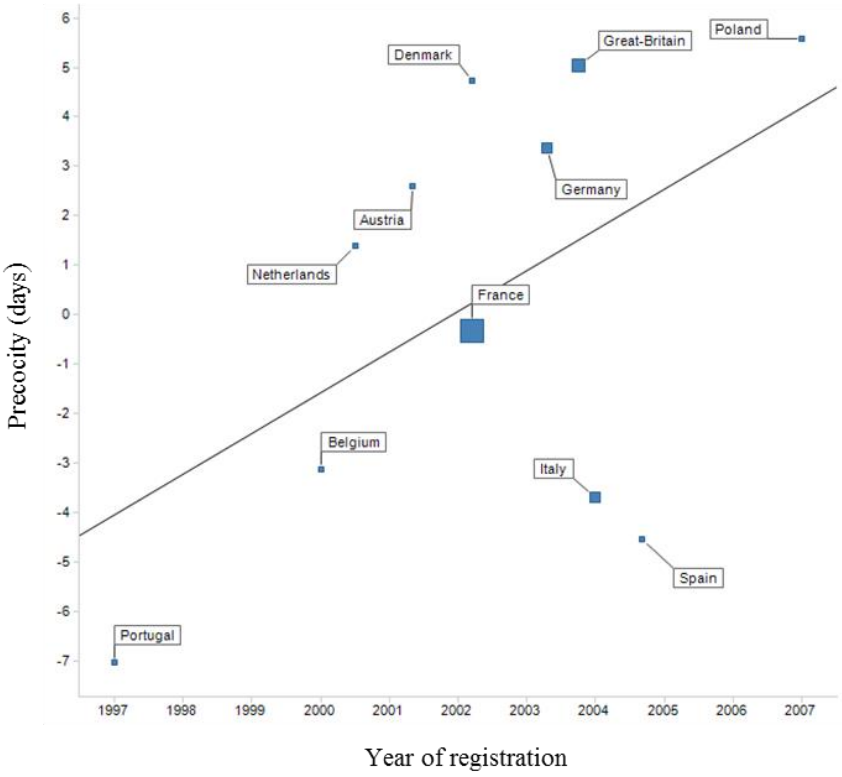
Supplementary data 4: Heritabilities at HN and LN, genetic correlations $R_{G(HN\_LN)}$ between HN and LN trials, and indirect selection efficiencies (ISE). ISE is computed as the efficiency of selecting in HN treatments to LN treatments (hgHN/ hgLN × $R_{G(HN\_LN)}$). Generalized heritabilities (h^2g) are calculated according to Cullis et al. (2006). varG and varε are respectively genetic and residual components of variances. Indirect selection is only efficient if the heritability is higher in the selecting environment than in the targeted one and exceeds the genetic correlation between these two environments. In this study that is never the case, and so indirect selection is never more efficient than direct selection. We conclude to direct selection at LN input is more efficient to target LN environments.

Trait	HN			LN			$R_{G(HN\_LN)}$	ISE
	h^2g	varG	varε	h^2g	varG	varε		
FLO	0.96	21.21	2.41	0.96	22.10	2.97	0.99	0.99
PH	0.88	35.05	11.21	0.86	31.06	13.60	0.93	0.94
SA	0.70	1836.47	1961.43	0.74	1431.98	1418.04	0.71	0.69
ADM_S	0.70	3.44×10^{5}	4.64×10^{5}	0.67	2.15×10^{5}	3.54×10^{5}	0.76	0.78
%N_S	0.58	1.88×10^{-3}	4.65×10^{-3}	0.64	1.03×10^{-3}	1.94×10^{-3}	0.67	0.63
GY	0.78	3.06×10^{5}	2.79×10^{5}	0.74	1.85×10^{5}	2.11×10^{5}	0.86	0.88
GPC	0.82	0.80	0.55	0.82	0.57	0.40	0.91	0.92
TKW	0.89	9.79	3.87	0.91	8.63	2.54	0.95	0.94
GNY	0.31	22.83	175.64	0.19	5.00	74.54	0.48	0.61
HI	0.69	3.91	5.74	0.76	4.83	4.84	0.84	0.80
NHI	0.41	1.42	6.85	0.39	1.03	5.53	0.49	0.51
NutE	0.75	8.19	8.71	0.75	20.85	22.39	0.86	0.86
GPD	0.62	0.23	0.46	0.63	0.19	0.37	0.73	0.72
NutE_Prot	0.78	2.30×10^{-5}	2.10×10^{-5}	0.76	3.06×10^{-5}	3.15×10^{-5}	0.81	0.82
NupE	0.18	3.26×10^{-4}	5.23×10^{-3}	0.18	3.33×10^{-4}	5.24×10^{-3}	0.26	0.25
NUE	0.74	5.53	6.32	0.74	8.40	9.83	0.78	0.78
NUE_Prot	0.76	1.18×10^{-5}	1.16×10^{-5}	0.81	2.42×10^{-5}	1.78×10^{-5}	0.88	0.86

Supplementary data 5: YR and Quality classes of 195 wheat varieties: Least Significant Difference (LSD) tests were performed to test whether the quality classes had different registration means using the "agricolae" package in R. Means with the same letter are not significantly different (P=0.05).

Quality classes	YR means	GPC
High quality	2003 a	9.95 b
Good quality	2003 a	9.81 b
Biscuit quality	2001 ab	9.79 b
Other uses	2001 ab	9.20 c
Very high quality	1999 b	12.11 a

Supplementary data 6: Precocity and origin of varieties used in this study. Precocity are calculated and centred as the days of flowering, once quality and height effects were removed. Origins are not homogeneously distributed among year of registration and are linked to precocity as wheat precocity conditions regional adaptation.



Supplementary data 7: Reduced height (Rht-1) gene frequencies in combination and effect on NUE. Rht-B1 and Rht-D1 genotyping data were available for 170 varieties out of the 195 used in the genetic progress study. The only Rht-B1b/Rht-D1b (double dwarf) cultivar was Courtot, but as it was registered before 1985 it was not included in the analyses. It resulted that only three *Rht-1* allelic combinations were present in our dataset. **A) Rht genes combinations description**. Taller varieties are older. The three combinations have been used in breeding at different periods. **B) Effect of combinations to NUE additive genetic effect.** These combinations had no effect on NUE additive genetic value when quality and precocity were already taken into account. **C) Decomposition of GxN interaction to NUE by ANOVA.** The Rht gene combination effect is confounded with the YR effect but explained more of the GxN interactions. Rht-D1b allele had the smallest GxN interaction to NUE at HN. **D) Boxplot of GxN interaction to NUE at HN for varieties registered in 2007 (8 Rht-B1b and 19 Rht-D1b)**. Difference was significant between Rht-B1b and Rht-D1b. Rht-D1b allele is indeed linked with the fact that recent varieties have GxN interactions which decrease their NUE at HN, and so increase their yield stability.

A)

Rht combinations	Number of varieties	Year of registration[1]	Plant height (cm)[1]
Rht-B1a/Rht-D1a (wild type)	20	1997 a	88.75 a
Rht-B1b/Rht-D1a (Rht1 type)	31	2001 b	79.39 b
Rht-B1a/Rht-D1b (Rht2 type)	119	2003 c	78.71 b

1. Tukey's test (P=0.05); means followed by a different letter are significantly different.

B)

	Adjusted r^2 (%)	Quality[1]	Precocity[1]	YR[1]	Rht Combinations[1]
With YR	62.4	45***	13***	5***	-
With *Rht-1* genes	57.6	45***	13***	-	1 ns.

1. Percentage of the variance explained by factors/variable (%)

*Fischer tests: ***, P-value <0.001; **, P-value <0.01; *, P-value <0.05 and ns., non-significant P-value>0.05*

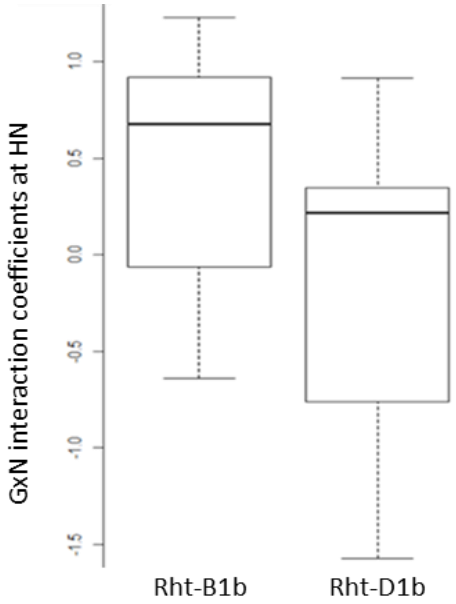
C)

	Adjusted r^2 (%)	Quality1	YR[1]	Rht Combinations[1]
With YR	3.9	6.01*	1.97*	-
With Rht -1 genes	6.6	6.01*	-	3.90*

1. Percentage of the variance explained by factors/variable (%)

*Fischer tests: ***, P-value <0.001; **, P-value <0.01; *, P-value <0.05 and ns., non-significant P-value>0.05*

D)



SUPPORTING INFORMATION ON PART III

[Supplementary data of Cormier et al. (2014) A genome-wide identification of chromosomal regions determining nitrogen use efficiency components in wheat (*Triticum aestivum* L.). Theor Appl Genet 127:2679-2693 **and** Cormier et al. (2015) Detection of natural variants of NAM-A1 in bread wheat. Submitted to Agronomy]

Supplementary data 1: Description of the experimental design where wheat genotypes were evaluated at high N level and low N level (from Cormier et al. 2013). NTA_{max} corresponds to the 95th percentile of total nitrogen per area at maturity for all the genotypes present in the trial and is an estimate of N available (soil + fertiliser N).

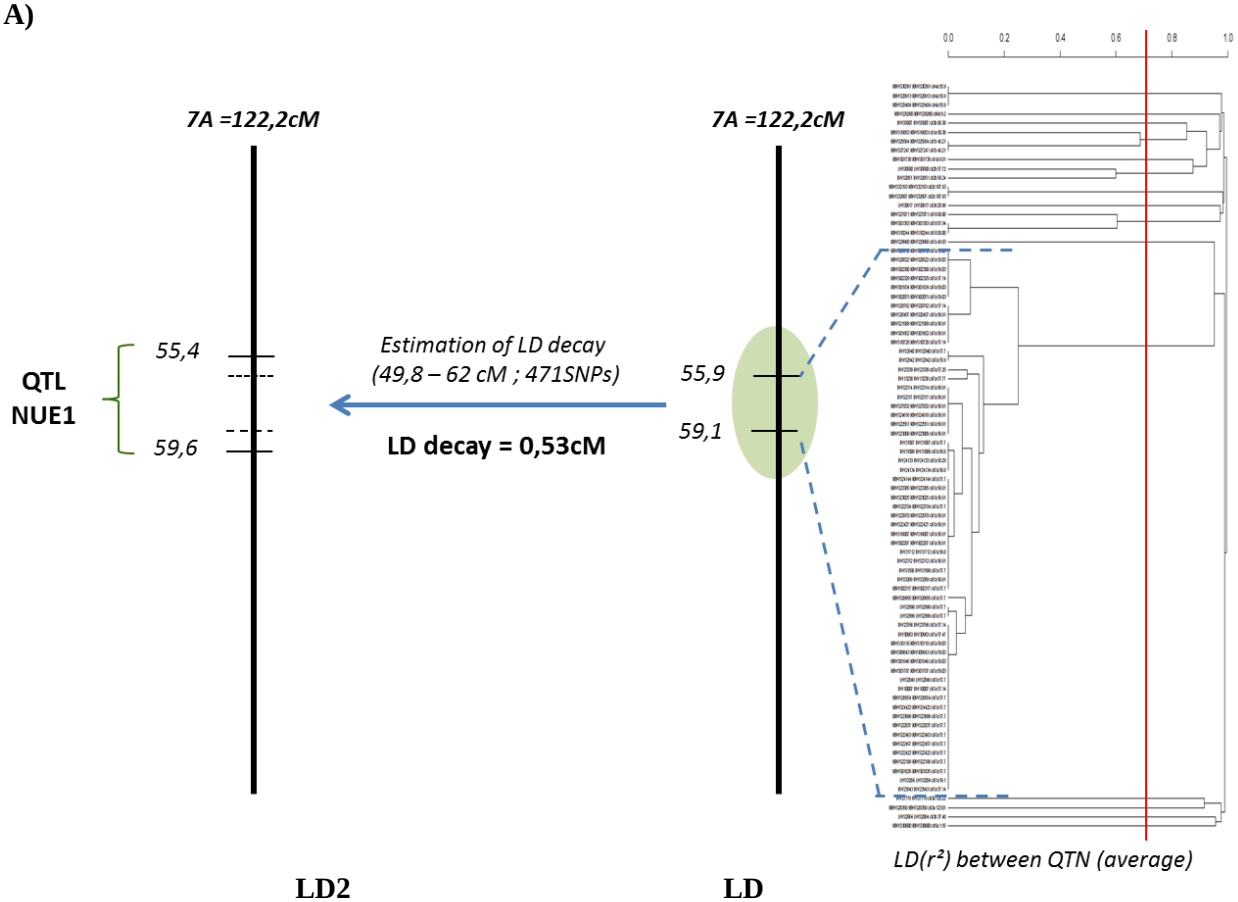
Site x Season	Season	Location	Soil type	Genotypes tested	Residual soil N (kg N ha^{-1})	N supply [1] (kg N ha^{-1})		NTA_{max} (kg N ha^{-1})	
						HN	LN	HN	LN
EM08	07/08	Estrées-Mons (49.8N,3.03E)	Clay loam	206*	67	50+70+50	0+70+0	206	144
EM09	08/09			208*	30	50+50+50	0+50+0	241	111
VB08	07/08	Villiers le Bacle (48.7N,2.1E)	Clay loam	197	106	0+66.5+60	0+44+0	242	157
VR09	08/09	Vraux (49.0N,4.2E)	White Chalk	196	30	60+100+60	60+60+0	236	173

1. *Nsupply: fertiliser supply at end of winter + at Z30 + at Z32.*

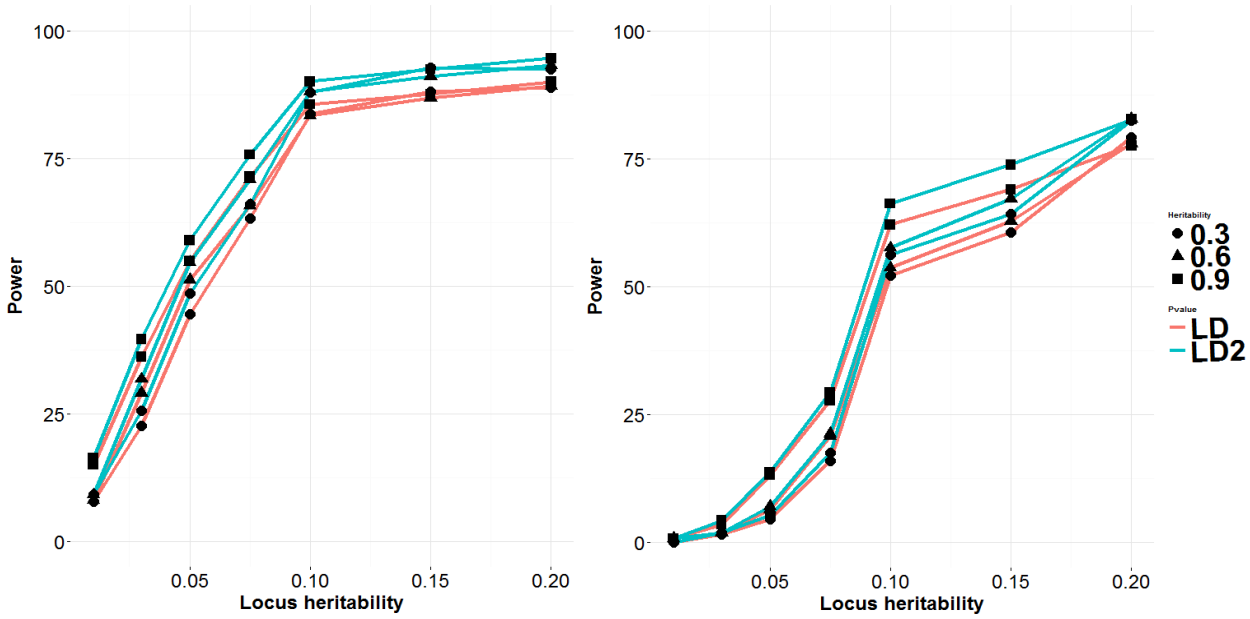
**controls: Apache, Orvantis, Caphorn, and Soissons (2007/08) or Premio (2008/09)*

Supplementary data 2: How did we define QTL from GWAS results? (A) Description of the method used to define QTL from GWAS results. The first step is based on LD between QTN (LD). A clustering by average distance between QTN was made with a cut-off = 1- "critical LD". The second (LD2) aimed to extend the first boundaries to take into account a possible LD with the causal mutation at the first boundaries. **(B) Influence of the extension of QTL boundaries (LD2) on the relation between locus heritability and power of detection in the association panel at a LOD score threshold of 3 (left) and 6 (right) for three narrow-sense heritabilities.** Power simulations were conducted as described in Mat & Meth. At a LOD score threshold of 3, the power increase average 4% when QTL were extended by using the LD2 steps and QTL size increase averaged 1.7 cM. **(C) Evolution of the false positive rate in function of locus and trait heritabilities, and LOD score threshold.** The false positive rate is defined as the proportion of chromosomal region which were defined but did not contain the causal mutation. QTL boundaries were computed following the two steps previously described (Supp data 1A). Power simulations were conducted as described in Mat & Meth section "Phenotype simulation and power". **(D) Influence of the extension of QTL boundaries (LD2) on false positive rate at a LOD score threshold of 3.** The false positive rate is defined as the proportion of chromosomal regions which were defined but did not contain the causal mutation. Power simulations were conducted as described in Mat & Meth section "Phenotype simulation and power".

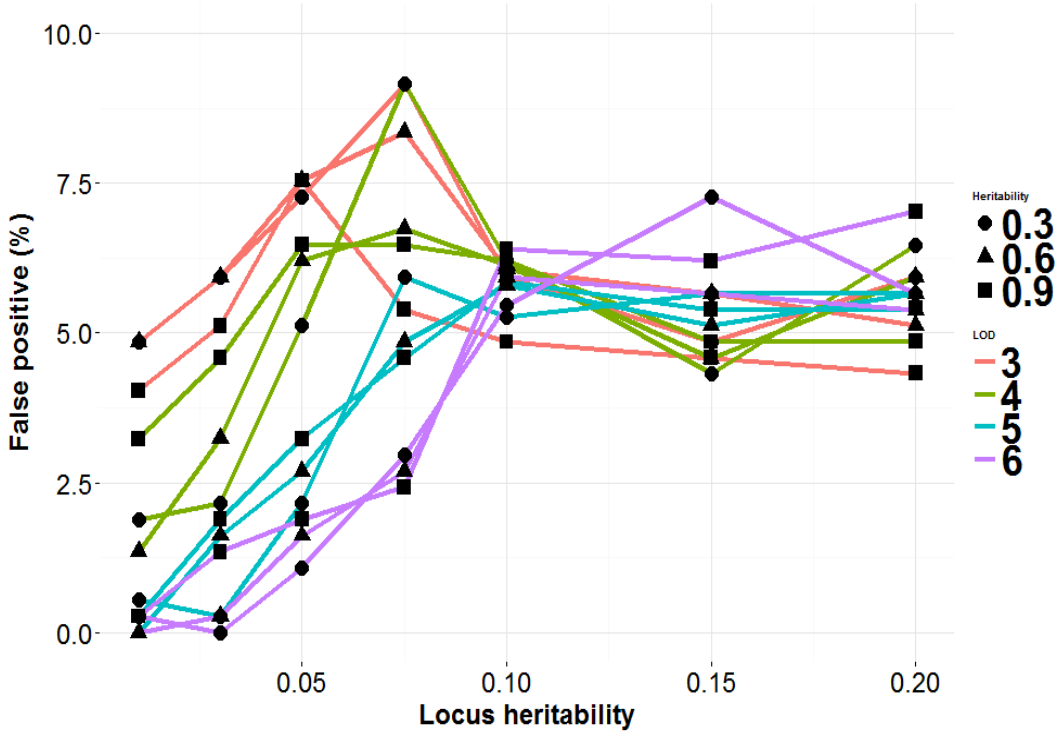
A)



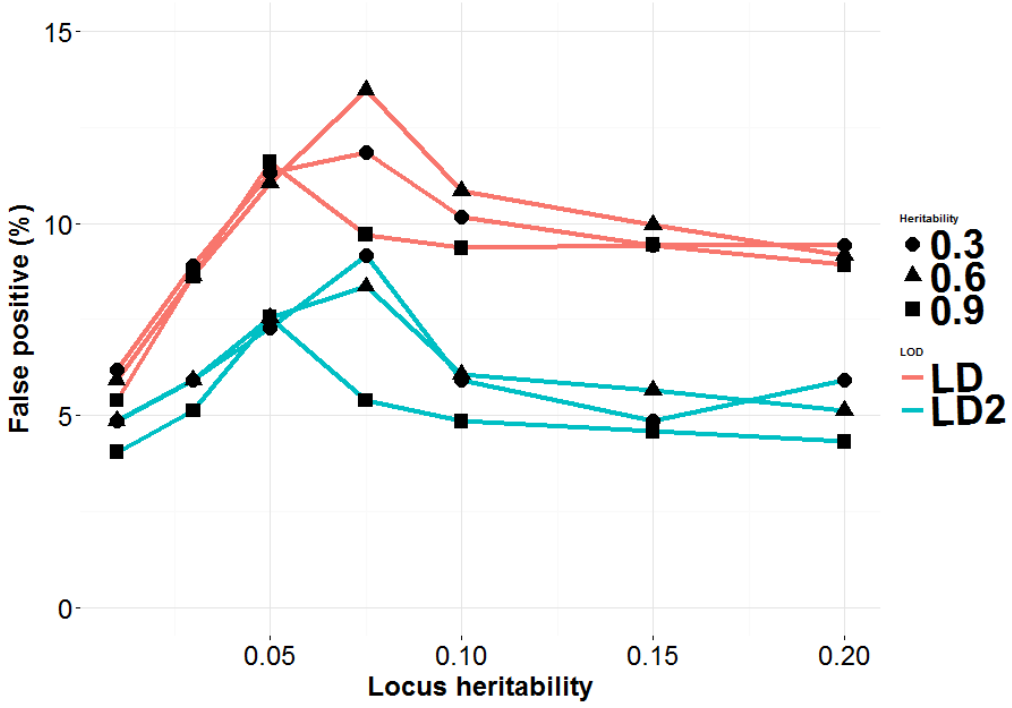
B)


Power
Locus heritability
Heritability
0.3
0.6
0.9
Pvalue
LD
LD2

C)


False positive (%)
Locus heritability
Heritability
0.3
0.6
0.9
LOD
3
4
5
6

D)


15
10
5
0
False positive (%)
0.05
0.10
0.15
0.20
Locus heritability
Heritability
0.3
0.6
0.9
LOD
LD
LD2

Supplementary data 3: Description of QTL. Minor allele frequency **(**MAF) and effect are the mean of significant SNP (QTN) within a QTL. LOD and r² are the max on significant SNP within a QTL. QTL boundaries are described by the closest markers on each side with a previously published map location.

Trait	QTL_name	MAF	Effect	LOD	r²	ch	From	To	Boundaries
NUE	NUE8	0.11	1.07	3.58	0.12	ch1a	0.00	25.37	GDM33-FBA393
NutE_Prot	NutE_Prot12	0.11	0.00	3.93	0.14	ch1a	0.00	25.37	GDM33-FBA393
HI	HI9	0.35	-0.65	3.26	0.02	ch1a	49.55	50.47	CFD65-GPW3083
EFFG	EFFG10	0.38	1.58	3.60	0.06	ch1a	55.79	56.25	BCD808A-WMC11
NHI	NHI11	0.20	-0.46	3.15	0.04	ch1a	55.79	56.25	BCD808A-WMC11
EFFREMN	EFFREMN8	0.10	1.06	3.44	0.06	ch1a	61.13	62.36	WPT-9757-BCD808B
ADM_FLO	ADM_FLO9	0.06	-339.27	3.17	0.07	ch1a	77.83	79.71	EDM80-GWM497
FLO	FLO16	0.09	1.77	3.48	0.10	ch1a	77.87	81.58	EDM80-WPT4658
EFFG	EFFG2	0.37	-1.53	3.04	0.06	ch1a	92.50	93.12	WPT1770-MWG632
%N_S	%N_S2	0.14	-0.02	6.35	0.21	ch1b	2.92	16.91	MGL77-WPT2230
NHI	NHI3	0.13	0.57	3.54	0.11	ch1b	3.54	16.48	KSUD14-FBA199
EFFREMN	EFFREMN4	0.11	1.00	3.39	0.07	ch1b	6.87	10.04	STM542ACAG-TPT5249
ADM_S	ADM_S5	0.21	195.66	3.95	0.11	ch1b	7.13	8.23	WPT3465-WPT1972
GNY	GNY4	0.08	2.78	3.75	0.06	ch1b	8.69	30.58	WPT1972-WMC419
REMN	REMN3	0.11	3.80	3.92	0.04	ch1b	8.69	10.04	WPT1972-TPT5249
ABSN	ABSN6	0.12	3.29	3.57	0.07	ch1b	17.90	28.88	KSUF43B-WPT0697
ADM_S	ADM_S8	0.12	-235.10	3.01	0.08	ch1b	17.90	18.58	KSUF43B-GWM264D
EFFG	EFFG9	0.11	-2.52	4.78	0.09	ch1b	17.90	28.88	KSUF43B-WPT0697
INN_FLO	INN_FLO5	0.19	-0.01	3.21	0.13	ch1b	17.90	18.58	KSUF43B-GWM264D
NHI	NHI7	0.24	0.44	3.68	0.08	ch1b	17.90	18.58	KSUF43B-GWM264D
REMN	REMN6	0.15	-3.01	3.47	0.07	ch1b	17.90	18.58	KSUF43B-GWM264D
%N_S	%N_S19	0.21	-0.01	3.97	0.11	ch1b	17.90	18.58	KSUF43B-GWM264D
EFFREMN	EFFREMN10	0.07	-1.20	3.59	0.09	ch1b	27.41	29.67	GPW4069-WMC500B
NTA	NTA3	0.07	3.18	3.18	0.04	ch1b	28.89	31.18	WPT0697-BCD1124
NupEMat	NupEMat6	0.07	0.02	3.20	0.05	ch1b	28.89	31.18	WPT0697-BCD1124
TKW	TKW4_9	0.09	-1.62	3.46	0.09	ch1b	29.42	40.06	WMC500B-CFD48
%N_S	%N_S4	0.08	0.02	4.17	0.08	ch1b	37.23	38.82	KU136-WPT5485
NSA	NSA1	0.06	1.54	3.16	0.05	ch1b	38.60	38.78	WPT1399-WPT5485
PH	PH14	0.10	2.19	3.39	0.05	ch1b	44.17	44.31	WPT0202-WPT0506
TKW	TKW5	0.41	0.28	3.50	0.09	ch1b	44.37	44.78	WPT0506-WPT0419
NupEMat	NupEMat5	0.45	-0.01	3.42	0.02	ch1b	59.79	60.71	DUPW214B-WMC430
HI	HI2	0.41	0.58	3.19	0.03	ch1b	88.55	88.74	GWM259C-WPT5164
ABSN	ABSN11	0.44	2.11	3.10	0.04	ch1b	91.67	91.85	WPT3950-CDO346
REMN	REMN12	0.44	-2.17	3.23	0.05	ch1b	91.67	91.85	WPT3950-CDO346
DMGY	DMGY9	0.05	252.67	3.14	0.07	ch1b	92.12	92.30	CDO346-CDO346
GNY	GNY8	0.32	1.55	3.28	0.07	ch1b	93.30	93.46	WPT1973-WPT1973
GNY	GNY6	0.14	-2.47	3.19	0.08	ch1b	94.28	94.43	KSUI27B-WPT3177
ADM_FLO	ADM_FLO8	0.15	-239.78	3.37	0.10	ch1d	51.24	56.66	WPT665814-WPT6316
GNY	GNY7	0.16	-2.25	4.42	0.10	ch1d	64.01	89.59	WPT8854-GPW300
NTA	NTA7	0.16	-2.61	4.32	0.11	ch1d	64.01	89.59	WPT8854-GPW300
NupEMat	NupEMat8	0.18	-0.01	3.57	0.09	ch1d	64.01	89.59	WPT8854-GPW300
GNY	GNY2	0.43	-1.78	5.29	0.14	ch2a	52.11	62.35	WMC326-GPW5257
FLO	FLO15	0.31	1.16	3.50	0.00	ch2a	54.26	57.04	CDO1090-GWM614
NTA	NTA2	0.44	-2.00	3.84	0.09	ch2a	54.68	58.95	GWM400-MRGA2
NupEMat	NupEMat3	0.44	-0.01	3.82	0.09	ch2a	54.68	58.95	GWM400-MRGA2
GPD	GPD2	0.45	-0.15	3.04	0.02	ch2a	56.17	58.95	GWM636-MRGA2
GPD	GPD6	0.33	0.12	3.06	0.06	ch2a	65.66	68.44	PSR332-WMC177
%N_S	%N_S21	0.16	0.02	3.00	0.10	ch2a	94.62	96.38	WMC522-WPT5251
%N_S	%N_S20	0.38	-0.01	3.01	0.11	ch2a	98.99	100.54	CFD55-GWM71D
NSA	NSA4	0.25	-1.19	3.73	0.02	ch2a	107.22	108.93	BQ161439-FBB353
%N_S	%N_S12	0.36	0.01	3.58	0.05	ch2a	120.19	120.82	GWM294-BCD1095
HI	HI16	0.49	0.18	3.40	0.04	ch2a	125.88	126.38	WMC261B-WPT1913
NupEFlo	NupEFlo2	0.42	-0.01	3.28	0.07	ch2a	139.35	140.35	WMC181C-WPT8326
ABSN	ABSN13	0.45	2.26	3.93	0.07	ch2a	140.05	142.22	WMC181C-WPT8326
EFFG	EFFG18	0.45	-1.62	3.62	0.07	ch2a	140.05	142.22	WMC181C-WPT8326
HI	HI17	0.29	-0.63	3.22	0.05	ch2a	174.26	176.47	CDO1410-BARC122
EFFG	EFFG8	0.08	-2.85	3.36	0.06	ch2a	203.71	204.02	WPT9302-WPT9302
NHI	NHI1	0.14	0.50	3.16	0.02	ch2a	206.68	208.11	WPT9302-WPT9302

Supplementary data 3 – continued

Trait	QTL_name	MAF	Effect	LOD	r^2	ch	From	To	Boundaries
NSA	NSA5	0.44	-0.01	3.07	0.06	ch2b	5.16	6.13	WMC661-WMC154A
INN_FLO	INN_FLO1	0.20	0.02	3.99	0.10	ch2b	8.01	8.99	WPT9859-WPT8970
NFA	NFA1_3	0.22	-2.56	3.30	0.06	ch2b	8.01	9.71	WPT9859-WPT8970
%N_FLO	%N_FLO2	0.24	0.03	4.28	0.10	ch2b	8.01	8.99	WPT9859-WPT8970
ABSN	ABSN5	0.23	2.54	3.59	0.05	ch2b	8.81	9.62	WPT8970-WPT8970
DMGY	DMGY7	0.06	-267.93	3.34	0.01	ch2b	8.81	9.62	WPT8970-WPT8970
EFFG	EFFG6	0.23	-1.86	3.63	0.06	ch2b	8.81	9.62	WPT8970-WPT8970
FLO	FLO7	0.46	0.99	3.47	0.07	ch2b	8.81	9.62	WPT8970-WPT8970
NupEFlo	NupEFlo1	0.21	-0.01	3.22	0.06	ch2b	8.81	9.71	WPT8970-WPT8970
REMN	REMN2	0.21	-2.52	3.58	0.06	ch2b	8.81	9.71	WPT8970-WPT8970
EFFG	EFFG20	0.23	-1.75	3.00	0.05	ch2b	10.70	11.48	GPW4016-WPT3592
PH	PH9	0.07	2.91	3.30	0.04	ch2b	27.45	28.48	WMC154D-WMC154D
%N_S	%N_S9	0.37	-0.01	4.91	0.07	ch2b	38.41	39.30	WPT4301-WPT1489
FLO	FLO14	0.47	1.17	3.99	0.15	ch2b	40.16	41.66	WPT9402-WPT5707
FLO	FLO6	0.12	1.52	3.57	0.09	ch2b	43.79	46.03	WPT6932-WMC770
NTA	NTA4	0.09	-3.12	3.01	0.11	ch2b	53.14	54.40	WPT6192-CFD11
NSA	NSA13	0.43	-0.85	3.16	0.06	ch2b	54.26	55.85	WPT1127-WPT2120
ADM_FLO	ADM_FLO2	0.33	207.46	3.99	0.04	ch2b	55.63	56.65	WPT2120-SHH293
FLO	FLO3	0.34	1.01	3.06	0.03	ch2b	55.63	56.65	WPT2120-SHH293
HI	HI3	0.05	1.32	3.01	0.13	ch2b	55.63	56.65	WPT2120-SHH293
INN_FLO	INN_FLO2	0.34	-0.01	3.25	0.06	ch2b	55.63	56.65	WPT2120-SHH293
%N_FLO	%N_FLO3	0.34	-0.03	3.71	0.05	ch2b	55.63	56.65	WPT2120-SHH293
DMGY	DMGY4	0.14	182.54	3.95	0.22	ch2b	55.72	58.23	WPT2120-ABC306
NUE	NUE4	0.14	1.03	3.59	0.21	ch2b	55.72	58.23	WPT2120-ABC306
GPC	GPC10	0.14	-0.28	3.07	0.20	ch2b	56.93	57.71	SHB123-GPW4354
NUE_Prot	NUE_Prot11	0.14	0.00	3.27	0.21	ch2b	56.93	57.71	SHB123-GPW4354
NutE	NutE6	0.14	1.42	3.02	0.16	ch2b	56.93	57.71	SHB123-GPW4354
PH	PH10	0.13	-2.62	4.29	0.20	ch2b	56.93	58.40	SHB123-ABC306
NutE_Prot	NutE_Prot5	0.13	0.00	3.16	0.20	ch2b	57.36	58.07	GPW7438-GPW4354
ADM_FLO	ADM_FLO3	0.41	-180.65	3.89	0.08	ch2b	62.84	64.23	BARC1064-WPT0709
EFFG	EFFG14	0.13	-2.29	3.83	0.07	ch2b	66.10	69.55	GPW7808-MWG660
NutE_Prot	NutE_Prot10	0.05	0.00	3.23	0.20	ch2b	67.17	69.56	BCD1119-MWG660
ADM_S	ADM_S16	0.39	176.01	3.11	0.03	ch2b	67.75	70.17	GWM129-GWM388
EFFREMN	EFFREMN6	0.23	-0.78	3.45	0.05	ch2b	67.75	70.17	GWM129-GWM388
HI	HI18	0.05	1.30	3.17	0.16	ch2b	67.75	70.17	GWM129-GWM388
NSA	NSA7	0.08	1.39	3.03	0.05	ch2b	67.75	70.17	GWM129-GWM388
GPD	GPD7	0.42	0.14	3.19	0.08	ch2b	68.93	71.36	GPW3050-BM134420
NUE_Prot	NUE_Prot10	0.41	0.00	3.31	0.08	ch2b	68.93	72.35	GPW3050-CNL6A
GPD	GPD5	0.15	-0.17	3.04	0.10	ch2b	75.60	78.10	WMC441-CFE52
SA	SA1	0.12	-17.56	3.66	0.12	ch2b	85.81	88.61	WMC360-WPT9190
NFA	NFA13	0.22	-2.52	3.08	0.07	ch2b	96.43	96.72	WPT2929-WPT2929
TKW	TKW8	0.28	0.88	3.08	0.01	ch2d	26.16	28.15	WPT6657-WMC111
ADM_S	ADM_S4_14	0.19	228.53	5.35	0.11	ch2d	45.31	52.97	GPW4321-WMC470
FLO	FLO13	0.23	-1.23	4.03	0.14	ch2d	45.31	52.32	GPW4321-WMC14
HI	HI20	0.39	-0.88	5.97	0.05	ch2d	45.31	52.32	GPW4321-WMC14
INN_FLO	INN_FLO3_8	0.19	-0.02	4.37	0.13	ch2d	45.31	52.97	GPW4321-WMC470
%N_FLO	%N_FLO4_9	0.19	-0.05	5.50	0.18	ch2d	45.31	52.97	GPW4321-WMC470
FLO	FLO4	0.19	1.97	7.90	0.25	ch2d	51.55	52.97	WMC14-WMC470
%N_S	%N_S6	0.19	-0.01	3.13	0.08	ch2d	51.55	52.97	WMC14-WMC470
ADM_S	ADM_S2	0.21	181.66	3.41	0.02	ch2d	64.87	70.44	CFD255-CFA2201
ADM_FLO	ADM_FLO13	0.24	191.31	3.05	0.06	ch2d	66.17	70.84	FBB279-CFA2201
NUE_Prot	NUE_Prot1	0.39	0.00	3.28	0.07	ch2d	67.75	74.51	GWM102-STM590TCAC
NutE	NutE1	0.40	0.98	3.09	0.08	ch2d	69.94	74.53	CFA2201-STM590TCAC
FLO	FLO8	0.23	1.12	3.13	0.01	ch2d	102.48	102.51	GPW308-GPW308
NSA	NSA3	0.47	-0.78	3.16	0.10	ch2d	104.26	104.50	WPT2781-WPT2781
NUE	NUE15	0.31	-0.57	3.23	0.01	ch2d	107.76	108.10	GPW5237-TAM8

Supplementary data 3 – continued

Trait	QTL_name	MAF	Effect	LOD	r^2	ch	From	To	Boundaries
%N_S	%N_S13	0.21	-0.02	4.40	0.05	ch3a	55.67	71.53	WMC388C-CDO281
ADM_FLO	ADM_FLO11	0.47	168.18	3.05	0.01	ch3a	57.05	57.79	WPT5766-BCD1823
FLO	FLO2	0.47	1.02	3.29	0.02	ch3a	57.05	57.79	WPT5766-BCD1823
NFA	NFA8_7	0.28	-2.51	3.56	0.08	ch3a	62.53	74.44	TPT1143-GWM638
NupEFlo	NupEFlo6_5	0.28	-0.01	3.96	0.08	ch3a	62.53	74.44	TPT1143-GWM638
HI	HI13	0.10	-0.84	3.11	0.01	ch3a	109.65	110.93	BARC51-WPT5125
GNY	GNY3	0.36	-1.69	3.74	0.07	ch3a	115.72	116.35	WPT9268-WMC169
DMGY	DMGY11	0.31	-120.40	4.09	0.01	ch3a	122.73	123.29	WPT1816-GWM666B
NUE	NUE10	0.31	-0.70	4.26	0.01	ch3a	122.73	123.29	WPT1816-GWM666B
NutE_Prot	NutE_Prot15	0.31	0.00	3.25	0.01	ch3a	122.73	123.29	WPT1816-GWM666B
ADM_S	ADM_S13	0.15	-192.35	3.46	0.06	ch3a	123.35	123.90	WPT1596-WPT2813
DMGY	DMGY6	0.44	106.76	4.15	0.02	ch3a	128.15	128.30	WPT6234-WPT6234
NUE	NUE6	0.44	0.67	4.78	0.03	ch3a	128.15	128.30	WPT6234-WPT6234
NutE_Prot	NutE_Prot8	0.44	0.00	3.75	0.02	ch3a	128.15	128.30	WPT6234-WPT6234
GPC	GPC4	0.36	0.18	3.44	-0.01	ch3a	131.80	132.01	CDO482-CDO482
NUE_Prot	NUE_Prot3	0.36	0.00	3.36	-0.01	ch3a	131.80	132.01	CDO482-CDO482
NutE	NutE2	0.36	-0.92	3.52	0.00	ch3a	131.80	132.01	CDO482-CDO482
SA	SA4	0.36	-10.83	3.16	0.10	ch3a	131.80	132.01	CDO482-CDO482
SA	SA10	0.13	15.31	3.82	0.04	ch3a	133.40	133.63	CDO482-CDO482
ADM_S	ADM_S9	0.12	-206.88	3.02	0.09	ch3b	27.84	28.31	WMM1344-WPT1336
EFFG	EFFG15	0.15	2.09	3.09	0.06	ch3b	27.84	28.31	WMM1344-WPT1336
NSA	NSA6	0.13	-1.14	3.42	0.04	ch3b	28.51	28.94	WPT1336-WPT1741
FLO	FLO20	0.43	-0.97	3.54	0.10	ch3b	36.43	36.75	CFB3023-CFB3023
%N_S	%N_S5	0.22	-0.01	3.12	0.07	ch3b	36.69	36.96	CFB3023-GPW3092
ADM_FLO	ADM_FLO5	0.27	-188.05	3.27	0.01	ch3b	37.37	37.58	WMM1441-WMM1441
NUE	NUE5	0.27	-0.71	3.33	0.01	ch3b	37.37	37.58	WMM1441-WMM1441
NutE_Prot	NutE_Prot6	0.27	0.00	3.23	0.00	ch3b	37.37	37.58	WMM1441-WMM1441
HI	HI4	0.48	0.57	3.39	0.05	ch3b	50.62	50.71	FBB24-FBB24
NSA	NSA8	0.06	1.55	3.36	0.08	ch3b	50.68	50.76	FBB24-FBB24
%N_S	%N_S11	0.06	0.02	3.27	0.04	ch3b	50.68	50.76	FBB24-FBB24
NFA	NFA10	0.21	-2.62	3.00	0.06	ch3b	50.85	50.94	FBB24-FBB24
GPC	GPC7	0.34	-0.21	3.09	0.13	ch3b	51.21	51.30	WMC540-WMC540
NUE_Prot	NUE_Prot7	0.36	0.00	3.07	0.12	ch3b	51.21	51.30	WMC540-WMC540
ABSN	ABSN8	0.09	-3.66	3.25	0.06	ch3b	51.27	51.36	WMC540-WMC540
HI	HI11	0.07	1.26	3.46	0.14	ch3b	51.98	52.06	CFP3112-CFP3112
TKW	TKW2	0.16	1.19	3.09	0.10	ch3b	52.16	52.24	CFB3260-CFB3260
HI	HI15	0.09	1.04	3.57	0.10	ch3b	88.11	88.64	CFB3440-CFB3440
NUE	NUE2	0.08	1.07	3.13	0.09	ch3b	88.11	88.64	CFB3440-CFB3440
PH	PH11	0.12	-2.67	4.39	0.13	ch3b	88.11	88.64	CFB3440-CFB3440
SA	SA9	0.15	-13.69	3.29	0.05	ch3b	91.45	92.06	WMM1133-WMM1133
SA	SA5	0.41	11.79	3.78	0.08	ch3b	101.30	101.61	CFE365-CFE365
EFFREMN	EFFREMN13	0.32	-0.67	3.30	0.05	ch3d	0.00	11.03	GPW7053-WPT742732
NSA	NSA14	0.32	0.82	3.38	0.02	ch3d	0.00	11.03	GPW7053-WPT742732
EFFREMN	EFFREMN11	0.07	1.25	3.26	0.07	ch3d	24.50	24.54	GPW4451-GPW4451
NUE	NUE3	0.34	0.65	3.29	0.00	ch3d	26.95	26.97	GDM128-GDM128
NutE_Prot	NutE_Prot4	0.34	0.00	3.27	0.00	ch3d	26.95	26.97	GDM128-GDM128
NFA	NFA5	0.17	-2.73	3.09	0.04	ch4a	49.95	50.81	GDM141-FBA147
NupEFlo	NupEFlo3	0.17	-0.02	3.65	0.05	ch4a	49.95	50.81	GDM141-FBA147
REMN	REMN5	0.17	-2.82	3.53	0.05	ch4a	49.95	50.81	GDM141-FBA147
SA	SA2	0.08	-25.63	5.00	0.11	ch4a	54.53	57.14	WPT7558-BCD8
HI	HI7	0.16	0.78	4.16	0.11	ch4a	54.70	55.51	WMC15-GPW4182
NUE	NUE11	0.26	0.71	3.15	0.11	ch4a	55.50	56.30	GPW4182-WMC757
NutE_Prot	NutE_Prot16	0.26	0.00	3.43	0.12	ch4a	55.50	56.30	GPW4182-WMC757
ADM_FLO	ADM_FLO6	0.18	239.90	3.22	0.12	ch4a	56.01	56.83	FBA211A-GWM610
NFA	NFA12	0.05	-4.92	3.08	0.08	ch4a	56.31	57.16	WMC757-GPW1010
PH	PH5	0.23	-1.81	3.76	0.08	ch4a	66.91	67.47	WPT0162-WPT3638
EFFG	EFFG4	0.37	-1.64	3.37	0.06	ch4a	67.40	67.96	WPT3638-WPT4660
EFFREMN	EFFREMN3	0.37	-0.61	3.00	0.08	ch4a	67.40	67.96	WPT3638-WPT4660
EFFG	EFFG19	0.30	-1.56	3.13	0.05	ch4a	71.23	72.68	CDO495-CD920298
NupEMat	NupEMat4	0.22	-0.01	3.06	0.06	ch4a	73.93	74.87	GWM397-GPW7020
TKW	TKW10	0.14	-1.11	3.09	0.07	ch4a	97.22	98.88	GPW2244-WPT2006
%N_S	%N_S3	0.09	-0.02	4.17	0.04	ch4a	115.42	115.66	SHH114-WPT9901
NSA	NSA2	0.08	-1.57	3.64	0.04	ch4a	115.45	115.57	SHH114-FBB154
GPC	GPC9	0.16	-0.23	3.30	0.04	ch4a	115.91	116.09	WPT5172-WPT2780
NUE_Prot	NUE_Prot9	0.16	0.00	3.14	0.04	ch4a	115.91	116.09	WPT5172-WPT2780
NutE	NutE4	0.16	1.18	3.06	0.05	ch4a	115.91	116.09	WPT5172-WPT2780
ABSN	ABSN1	0.10	3.51	3.32	0.06	ch4a	121.59	121.79	WMC497-WMC722

Supplementary data 3 – continued

Trait	QTL_name	MAF	Effect	LOD	r^2	ch	From	To	Boundaries
ADM_S	ADM_S11	0.45	-139.59	3.06	0.04	ch4b	-1.32	0.08	BE637594-BE637594
HI	HI19	0.45	0.60	3.87	0.14	ch4b	-1.32	0.08	BE637594-BE637594
%N_S	%N_S10	0.16	0.01	3.13	0.12	ch4b	33.89	44.35	PSP3163-WMC657
NHI	NHI5	0.41	0.39	3.27	0.04	ch4b	53.41	54.39	GPW4075-SHI211
GNY	GNY11	0.40	1.50	3.11	0.03	ch4b	62.62	66.38	GWM573-WPT8756
ABSN	ABSN14	0.13	3.11	3.30	0.06	ch4b	77.32	83.24	WPT3917-WPT5996
ADM_S	ADM_S17	0.46	-199.24	3.47	0.00	ch4d	21.07	25.93	CFD18-WPT0941
PH	PH17	0.46	-2.01	3.47	0.05	ch4d	21.07	25.93	CFD18-WPT0941
%N_S	%N_S7	0.07	-0.02	3.33	0.05	ch4d	31.20	35.26	GBXG102-BLT101
ADM_S	ADM_S6	0.26	-156.72	3.80	0.03	ch5a	-0.18	2.62	GPW4432-WPT2768
HI	HI8	0.26	0.58	3.24	0.06	ch5a	-0.18	2.33	GPW4432-GWM241
PH	PH6	0.26	-1.68	4.80	0.08	ch5a	-0.18	2.62	GPW4432-WPT2768
DMGY	DMGY12	0.20	-151.19	3.82	0.02	ch5a	1.09	2.00	GWM241-GWM241
NUE	NUE14	0.20	-0.79	3.35	0.02	ch5a	1.09	2.00	GWM241-GWM241
NutE_Prot	NutE_Prot17	0.20	0.00	3.46	0.02	ch5a	1.09	2.00	GWM241-GWM241
EFFREMN	EFFREMN2	0.22	0.82	4.37	0.08	ch5a	48.83	50.57	PSY-GPW3049
NupEMat	NupEMat1	0.16	0.01	3.44	0.08	ch5a	56.13	57.16	TPT9702-WPT0605
NHI	NHI8	0.49	0.51	3.56	0.03	ch5a	59.11	59.89	DOFA-DOFA
NutE_Prot	NutE_Prot3	0.30	0.00	3.45	0.01	ch5a	61.47	62.13	BCD926-GWM186
FLO	FLO9	0.24	1.07	3.11	0.05	ch5a	64.51	66.97	WG564-GWM96
GPC	GPC6	0.05	-0.44	3.92	0.23	ch5a	64.51	64.97	WG564-PSB85
NUE_Prot	NUE_Prot5	0.05	0.00	4.22	0.23	ch5a	64.51	64.97	WG564-PSB85
NutE_Prot	NutE_Prot14	0.05	0.00	3.52	0.28	ch5a	64.51	64.97	WG564-PSB85
FLO	FLO19	0.46	0.93	3.26	0.06	ch5a	69.92	70.84	MGB174-BCD1355
ABSN	ABSN9	0.18	-2.73	3.53	0.07	ch5a	70.67	71.77	BCD1355-FBB2
FLO	FLO18	0.25	-1.10	3.23	0.06	ch5a	71.09	72.19	BCD1355-BARC330
GNY	GNY9	0.21	-1.94	3.48	0.08	ch5a	133.21	133.52	ABG366-ABG366
NTA	NTA6	0.20	-2.27	3.45	0.10	ch5a	133.21	133.52	ABG366-ABG366
NupEMat	NupEMat11	0.19	-0.01	3.01	0.08	ch5a	133.21	133.52	ABG366-ABG366
FLO	FLO11	0.06	2.19	3.08	0.11	ch5a	143.36	143.81	WPT5096-WPT5096
ADM_S	ADM_S3	0.37	-151.96	3.01	0.03	ch5a	144.22	175.98	WPT5096-B1
NSA	NSA10	0.22	-0.87	3.05	0.08	ch5a	145.21	146.85	GWM595-GWM595
NHI	NHI6	0.27	0.44	3.55	0.07	ch5a	145.24	146.68	GWM595-GWM595
PH	PH16	0.24	-1.72	3.00	0.03	ch5a	146.88	148.78	GWM595-WMC524
NFA	NFA2	0.26	2.07	3.02	0.06	ch5a	147.11	148.89	WMC524-WMC524
EFFG	EFFG7	0.30	1.65	3.00	0.04	ch5a	149.51	151.21	WMC727-WMC727
PH	PH1_13	0.19	-2.44	4.25	0.10	ch5b	98.69	153.83	GWM540-WPT4577
NupEMat	NupEMat10	0.47	0.01	3.08	0.02	ch5b	98.94	107.23	FBA342-GBXG198
INN_FLO	INN_FLO6	0.37	-0.01	3.37	0.13	ch5b	103.16	121.42	GWM67-BCD351
GPC	GPC2	0.05	-0.48	3.27	0.27	ch5b	108.04	132.41	WPT6726-DUPW395
NUE_Prot	NUE_Prot2	0.05	0.00	3.18	0.26	ch5b	108.04	132.41	WPT6726-DUPW395
ABSN	ABSN10	0.15	3.03	3.86	0.06	ch5b	141.25	155.40	WMC289-CFD156
EFFG	EFFG12	0.15	-2.26	4.49	0.08	ch5b	141.25	142.85	WMC289-WMC289
NFA	NFA11	0.16	-3.51	4.43	0.10	ch5b	154.38	154.81	WPT2707-WPT2707
NupEFlo	NupEFlo8	0.16	-0.02	5.47	0.13	ch5b	154.38	154.81	WPT2707-WPT2707
REMN	REMN10	0.16	-3.12	3.70	0.08	ch5b	154.38	154.81	WPT2707-WPT2707
HI	HI12_14	0.08	1.21	5.29	0.12	ch5b	166.09	173.87	WPT8414-GDM116
NHI	NHI9	0.39	0.44	3.48	0.07	ch5b	166.09	171.62	WPT8414-WPT0517
INN_FLO	INN_FLO7	0.40	-0.01	3.57	0.19	ch5b	166.41	170.15	WPT8414-CFA2121B
%N_FLO	%N_FLO8	0.40	-0.03	3.80	0.26	ch5b	166.41	170.15	WPT8414-CFA2121B
TKW	TKW7	0.45	-0.83	3.10	0.07	ch5b	166.92	170.63	CDO584-WPT0517
NUE_Prot	NUE_Prot6	0.08	0.00	3.05	0.04	ch5b	171.48	173.87	WPT0517-GDM116
ADM_S	ADM_S1	0.27	186.27	3.50	0.11	ch5b	173.55	175.58	GDM116-WPT6880
%N_S	%N_S1	0.27	-0.01	3.15	0.05	ch5b	173.55	175.58	GDM116-WPT6880
NSA	NSA9	0.15	1.10	3.56	0.12	ch5b	195.63	195.75	TPT3144-WMC783
NutE	NutE3	0.15	-1.34	3.61	0.08	ch5b	195.63	195.75	TPT3144-WMC783
%N_S	%N_S14	0.15	0.02	4.42	0.14	ch5b	195.63	195.75	TPT3144-WMC783
%N_S	%N_S15	0.14	-0.02	3.82	0.03	ch5b	208.41	210.32	SSIB-PSR580

Supplementary data 3 – continued

Trait	QTL_name	MAF	Effect	LOD	r^2	ch	From	To	Boundaries
REMN	REMN9	0.31	-2.21	3.09	0.06	ch6a	3.71	4.79	WPT5395-WPT4752
GPD	GPD8	0.48	0.12	3.06	0.11	ch6a	8.02	8.90	WPT1377-WPT1377
%N_FLO	%N_FLO10	0.46	-0.03	3.00	0.09	ch6a	8.29	9.30	WPT1377-WPT730591
DMGY	DMGY3	0.09	186.21	3.03	0.17	ch6a	13.80	15.69	PTAG53-WPT0562
SA	SA3	0.13	14.76	3.09	0.04	ch6a	21.60	23.14	WPT671799-WPT3965
%N_S	%N_S18	0.06	-0.03	3.35	0.13	ch6a	25.96	26.47	WPT3091-WPT3091
%N_FLO	%N_FLO5	0.49	0.03	3.09	0.01	ch6a	27.79	28.35	PSR312-BARC118
PH	PH2	0.15	-2.33	3.86	0.08	ch6a	28.97	53.22	CFE80-GWM570
GNY	GNY5	0.27	2.04	3.93	0.07	ch6a	29.42	30.06	CFE80-GPW7455
HI	HI6	0.15	0.89	3.05	0.03	ch6a	30.45	31.09	GPW7455-BARC107
EFFREMN	EFFREMN7	0.32	0.66	3.24	0.07	ch6a	52.56	52.64	GPW3251-GPW3251
GPD	GPD1	0.17	-0.20	4.10	0.09	ch6a	52.56	52.64	GPW3251-GPW3251
%N_S	%N_S8	0.30	0.01	3.30	0.03	ch6a	52.56	52.64	GPW3251-GPW3251
DMGY	DMGY2	0.11	190.39	3.38	0.18	ch6a	52.67	52.75	GPW3251-GPW3251
NutE_Prot	NutE_Prot1	0.11	0.00	3.05	0.19	ch6a	52.67	52.75	GPW3251-GPW3251
ABSN	ABSN7	0.22	2.41	3.18	0.05	ch6a	58.11	58.19	CSB112-CSB112
EFFG	EFFG11	0.22	-1.95	3.70	0.06	ch6a	58.11	58.19	CSB112-CSB112
INN_FLO	INN_FLO4	0.23	-0.01	3.48	0.07	ch6a	85.41	87.10	GWM169-GPW5125
%N_FLO	%N_FLO6	0.23	-0.03	3.08	0.08	ch6a	85.41	87.10	GWM169-GPW5125
SA	SA8	0.12	17.91	3.96	0.07	ch6a	88.87	89.45	FBB70-GPW7388
TKW	TKW6	0.33	-0.87	4.12	0.10	ch6a	92.40	96.73	WPT0938-TPT4178
FLO	FLO12	0.20	1.21	3.06	0.08	ch6a	93.99	94.87	WPT0696-WPT9474
NTA	NTA8	0.14	-2.48	3.02	0.09	ch6a	93.99	94.87	WPT0696-WPT9474
NupEMat	NupEMat9	0.14	-0.01	3.30	0.09	ch6a	93.99	94.87	WPT0696-WPT9474
PH	PH12	0.08	-2.90	3.85	0.15	ch6a	94.85	96.34	WPT9474-WMC642
TKW	TKW3	0.29	0.99	4.12	0.12	ch6a	95.04	96.48	GWM427-TPT4178
ADM_S	ADM_S12	0.16	-204.79	3.00	0.06	ch6a	95.12	96.48	GWM427-TPT4178
NTA	NTA9	0.48	-1.82	3.19	0.05	ch6a	95.12	96.48	GWM427-TPT4178
ABSN	ABSN3	0.13	-3.26	3.59	0.07	ch6b	36.30	36.39	WPT4415-WPT4415
EFFG	EFFG5	0.13	2.30	3.30	0.06	ch6b	36.30	36.39	WPT4415-WPT4415
SA	SA11	0.06	-22.79	3.24	0.07	ch6b	36.49	36.58	WPT8721-WPT8721
ADM_FLO	ADM_FLO4	0.09	-294.15	3.51	0.04	ch6b	36.77	36.86	WPT5461-WPT5461
NFA	NFA6	0.17	-2.91	3.02	0.04	ch6b	64.22	65.54	SHI330-FBB130
NupEFlo	NupEFlo4	0.17	-0.02	3.22	0.04	ch6b	64.22	65.54	SHI330-FBB130
%N_FLO	%N_FLO7	0.13	-0.04	3.15	0.11	ch6b	64.90	66.13	SHI330-FBB130
NutE_Prot	NutE_Prot7_13	0.15	0.00	3.62	0.20	ch6d	8.39	9.95	WPT1519-WPT672044
GPC	GPC8	0.09	-0.36	4.47	0.20	ch6d	8.46	9.95	WPT1519-WPT672044
GPD	GPD4	0.09	-0.23	3.77	0.13	ch6d	8.46	9.95	WPT1519-WPT672044
NUE	NUE12	0.09	1.07	3.07	0.18	ch6d	8.46	9.95	WPT1519-WPT672044
NUE_Prot	NUE_Prot8	0.09	0.00	4.48	0.20	ch6d	8.46	9.95	WPT1519-WPT672044
NutE	NutE5	0.09	1.66	3.70	0.16	ch6d	8.46	9.95	WPT1519-WPT672044
SA	SA12	0.06	24.53	3.25	0.03	ch6d	8.46	9.95	WPT1519-WPT672044
FLO	FLO10	0.22	-1.14	3.19	0.03	ch6d	125.31	127.35	GPW5179-GPW5179
PH	PH3	0.40	-1.84	5.70	0.12	ch7a	4.87	7.55	WPT6034-WPT4835
NutE_Prot	NutE_Prot2	0.48	0.00	3.10	0.04	ch7a	9.55	10.71	WPT2903-WPT4126
NUE	NUE13	0.34	-0.64	3.03	0.13	ch7a	47.64	52.22	BARC222-WPT8897
NUE	NUE1	0.08	-1.34	4.27	0.12	ch7a	55.37	59.63	BARC174-GWM631
EFFG	EFFG17	0.10	-2.75	3.76	0.07	ch7a	65.66	65.74	WMC488-WMC488
GNY	GNY1	0.38	-1.51	3.49	0.05	ch7a	65.66	74.97	WMC488-WPT2083
NTA	NTA1	0.38	-1.74	3.33	0.03	ch7a	65.66	74.97	WMC488-WPT2083
NupEFlo	NupEFlo9	0.10	-0.02	3.08	0.05	ch7a	65.66	65.74	WMC488-WMC488
NupEMat	NupEMat2	0.38	-0.01	3.28	0.03	ch7a	65.66	74.97	WMC488-WPT2083
REMN	REMN4	0.13	3.18	3.76	0.09	ch7a	65.66	65.74	WMC488-WMC488
NHI	NHI10	0.45	0.34	3.18	0.01	ch7a	68.66	69.47	DUPW226-DUPW226
NSA	NSA15	0.45	-0.69	3.54	0.03	ch7a	68.66	69.47	DUPW226-DUPW226
NHI	NHI4	0.15	-0.54	3.17	0.05	ch7a	68.88	68.99	DUPW226-DUPW226
ADM_FLO	ADM_FLO12	0.18	216.95	3.72	0.12	ch7a	69.03	69.68	DUPW226-DUPW226
FLO	FLO5	0.30	1.11	3.82	0.02	ch7a	72.63	72.76	SALA-SALA
GPD	GPD3	0.38	-0.11	3.15	0.03	ch7a	74.87	74.97	WPT4665-WPT2083
ADM_FLO	ADM_FLO7	0.25	179.01	3.03	0.08	ch7a	75.88	75.98	TPT9518-TPT9518
REMN	REMN7	0.13	-3.24	3.60	0.06	ch7a	78.69	78.78	FBA350-FBA350
ADM_S	ADM_S7	0.44	165.27	3.55	0.04	ch7a	81.45	81.69	WMC346-WPT1424
HI	HI10	0.44	-0.56	3.00	0.00	ch7a	81.45	81.69	WMC346-WPT1424
PH	PH4	0.22	-2.09	4.45	0.17	ch7a	100.52	103.90	WMC809-WMC809

Supplementary data 3 – continued

Trait	QTL_name	MAF	Effect	LOD	r^2	ch	From	To	Boundaries
NutE_Prot	NutE_Prot18	0.26	0.00	3.03	0.06	ch7b	-1.59	2.60	WMC606-WMC323
GPC	GPC5	0.35	-0.20	3.03	0.12	ch7b	47.57	54.74	BE499017-WMC546C
NUE_Prot	NUE_Prot4	0.35	0.00	3.26	0.12	ch7b	47.57	54.74	BE499017-WMC546C
PH	PH15	0.17	-2.05	3.57	0.13	ch7b	47.57	54.74	BE499017-WMC546C
ADM_FLO	ADM_FLO10	0.17	-223.60	3.21	0.04	ch7b	51.58	60.31	WMC546C-WPT8849
SA	SA7	0.08	-21.04	3.10	0.02	ch7b	89.03	94.38	WPT8106-WPT1149
EFFREMN	EFFREMN9	0.06	1.39	3.32	0.08	ch7b	90.93	94.91	WPT8890-WPT4230
TKW	TKW1	0.09	-1.56	4.40	0.09	ch7b	95.07	101.43	WPT4230-BARC315
NSA	NSA11	0.17	1.03	3.15	0.11	ch7b	111.33	112.67	GPW4471-FBB352
REMN	REMN11	0.09	3.33	3.05	0.06	ch7b	111.33	112.67	GPW4471-FBB352
HI	HI1	0.19	0.72	3.51	0.13	ch7b	112.58	114.07	FBB352-GPW4369
ABSN	ABSN4	0.14	-2.98	3.64	0.07	ch7b	114.35	155.41	GPW4369-WPT8938
REMN	REMN8	0.16	2.83	3.26	0.08	ch7b	122.79	137.38	WPT3723-WPT5892
EFFG	EFFG3	0.15	2.13	3.60	0.07	ch7b	123.08	158.95	WPT3723-WPT5747
GNY	GNY10	0.38	-1.57	3.02	0.07	ch7b	144.16	147.99	WPT5463-STM5TCACA
ABSN	ABSN12	0.29	2.53	3.84	0.05	ch7b	151.50	152.06	DUPW398-BARC258
EFFG	EFFG16	0.29	-1.87	3.83	0.05	ch7b	151.50	152.06	DUPW398-BARC258
EFFREMN	EFFREMN5	0.26	0.76	3.89	0.09	ch7b	161.63	162.33	WPT9813-WPT1196
%N_S	%N_S16	0.20	-0.02	5.00	0.10	ch7b	162.83	166.23	WPT3530-WPT7113
DMGY	DMGY1	0.36	114.98	3.10	0.15	ch7b	166.11	166.85	WPT7113-BARC182
%N_S	%N_S17	0.22	-0.01	3.45	0.09	ch7b	166.85	167.59	BARC182-BARC97B
EFFREMN	EFFREMN12	0.08	1.21	3.00	0.06	ch7b	167.47	168.22	BARC97B-KSUE18B
ADM_S	ADM_S15	0.24	196.84	3.03	0.08	ch7b	182.26	185.22	AWM449-AWM449
NFA	NFA4	0.35	-2.30	3.37	0.05	ch7d	86.06	86.14	BARC352-BARC352
NUE	NUE7	0.15	0.91	3.58	0.06	ch7d	87.91	88.11	GPW334-GPW334
NutE_Prot	NutE_Prot11	0.15	0.00	4.49	0.09	ch7d	87.91	88.11	GPW334-GPW334
ADM_S	ADM_S10	0.06	296.19	3.34	0.07	ch7d	88.04	88.11	GPW334-GPW334
DMGY	DMGY10	0.19	160.50	3.82	0.07	ch7d	88.04	88.11	GPW334-GPW334
EFFG	EFFG1	0.34	1.63	3.17	0.05	ch7d	94.35	94.39	WPT4555-WPT4555
REMN	REMN1	0.34	2.49	4.07	0.08	ch7d	94.35	94.39	WPT4555-WPT4555

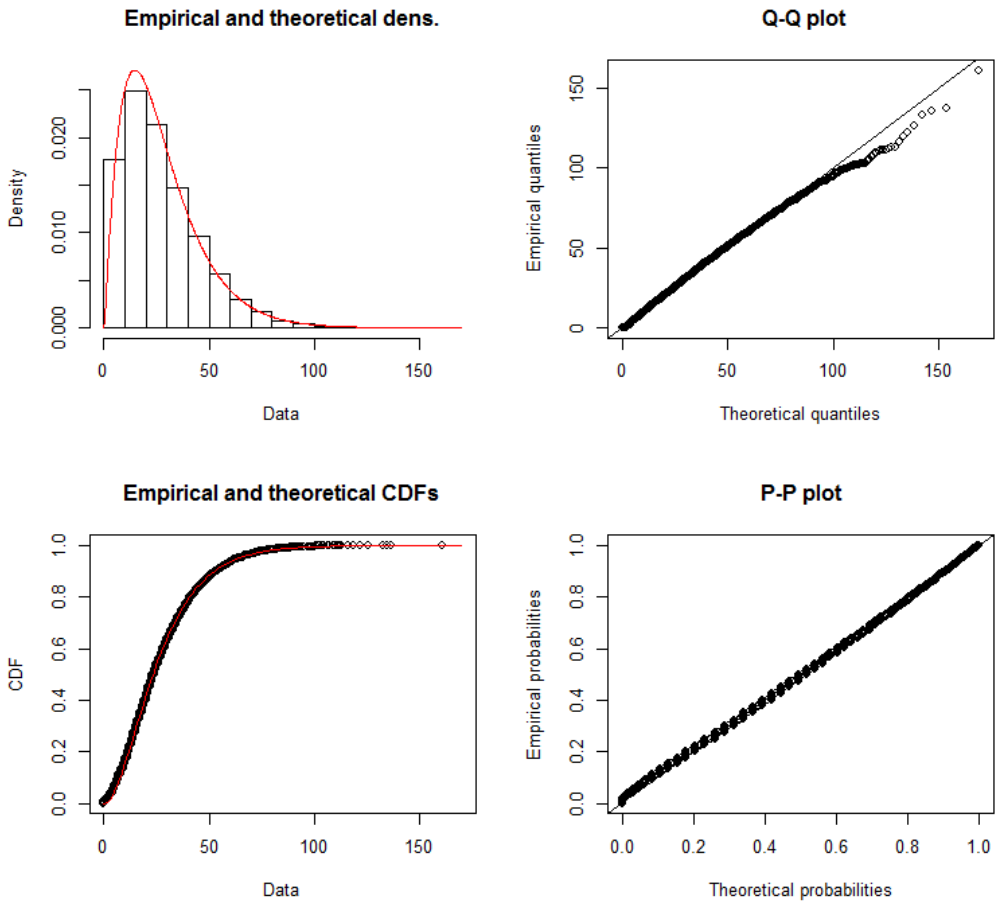
Supplementary data 4: Number of common QTLs between two traits. Numbers of common QTLs with opposite effects on traits are located in the inferior diagonal, same sign effect are in the superior diagonal, and the total number of QTL are on the diagonal.

	ABSN	ADM_FLO	ADM_S	DMGY	EFFG	EFFREMN	FLO	GNY	GPC	GPD	HI	INN_FLO	NFA	NHI	NSA	NTA	NUE	NUE_Prot	NupEFlo	NupEMat	NutE	NutE_Prot	PH	REMN	SA	TKW	X,N_FLO	X,N_S
ABSN	13				1			1				1		1														
ADM_FLO		12	1				2		1									1										
ADM_S			16		1		1				2	1			2								4					1
DMGY				10			1				1						5											
EFFG	-8		-2		19			1				1												4				
EFFREMN						12																						1
FLO		-1	-1	-1			18				2	2															2	2
GNY					-1		-2	11		2				3		5			1	5				1				
GPC									8			1						8					2					
GPD										8						2		2		2		1						
HI			-4	-1			-1				18	1		1	3			1									1	
INN_FLO	-1		-1		-1		-2				-1	7						1								1	5	
NFA													10						5					3			1	1
NHI			-1		-2							-2		10										1				
NSA						-2					-1				14	1												
NTA							-3							-1		8			1	6								
NUE																	14				2							
NUE_Prot		-1									-1							11					2					
NupEFlo																			7	1				3				1
NupEMat							-3					-1		-1						10			1					
NutE									-4	-1								-5			6				2			
NutE_Prot				-5						-1							-9				-2	16						
PH	-1							-1			-3									-1			14					
REMN	-3				-1			-1						-1		-1			-1	-1				12				1
SA						-1																			11			
TKW																-2										8	1	
%N_FLO			-1	-2			-2				-1		-1														8	
%N_S		-1	-3			-2	-2	-1						-2										-1				21

Supplementary data 5: Frequencies of colocalisation between traits underlying the colocalisation network. Results are read by row (example: all GPC QTL are also NUE_Prot QTL, but only 73% (8/11) of NUE_Prot QTL are GPC QTL).

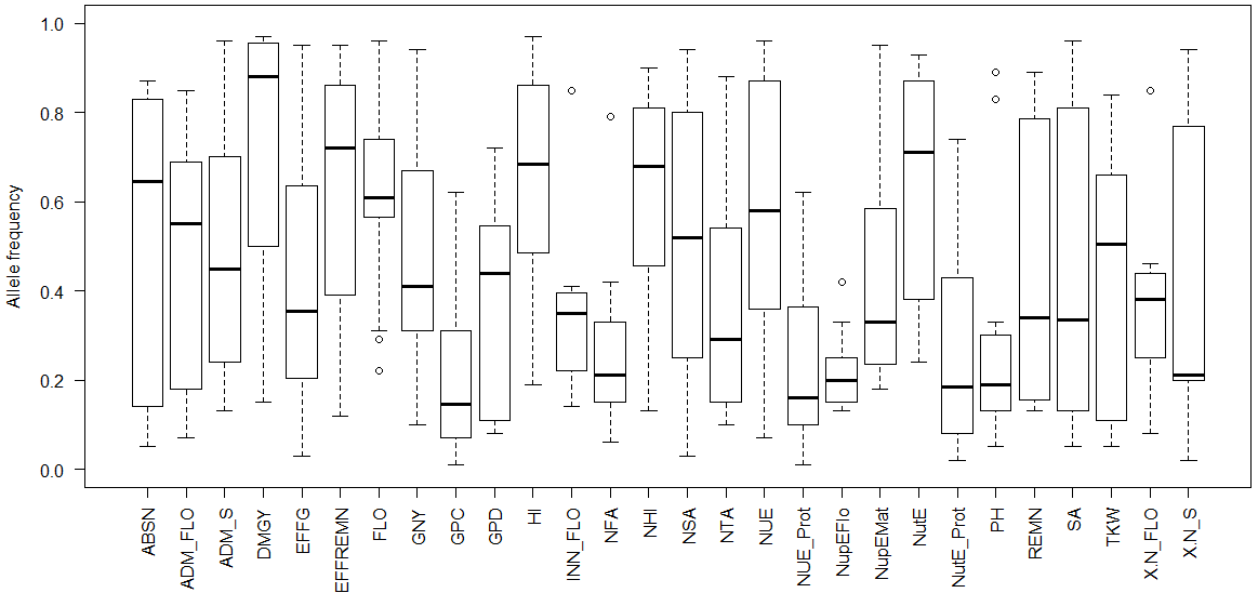
	ABSN	ADM_FLO	ADM_S	DMGY	EFFG	EFFREMN	FLO	GNY	GPC	GPD	HI	INN_FLO	NFA	NHI	NSA	NTA	NUE	NUE_Prot	NupEFlo	NupEMat	NutE	NutE_Prot	PH	REMN	SA	TKW	%N_FLO	%N_S
ABSN					69%			8%				15%		8%									8%	23%				
ADM_FLO			8%				25%		8%									17%										8%
ADM_S		6%			19%		13%				38%	13%		6%	13%								25%				6%	25%
DMGY							20%				20%						50%					50%					20%	
EFFG	47%		16%					11%				11%		11%										26%				
EFFREMN															17%										8%			25%
FLO		17%	11%	11%				11%			17%	22%				17%				17%							22%	22%
GNY	9%				18%		18%			18%				27%		45%			9%	45%			9%	18%				9%
GPC		13%										13%						100%			50%		25%					
GPD								25%								25%		25%		25%	13%	25%						
HI			33%	11%			17%					11%		6%	22%			11%					17%				11%	
INN_FLO	29%		29%		29%		57%		14%		29%			29%				14%		14%						14%	71%	
NFA																			50%					30%			20%	10%
NHI	10%		10%		20%			30%			10%	20%				10%				10%				20%				20%
NSA			14%			14%					29%					7%												
NTA							38%	63%		25%				13%	13%				13%	75%				13%		25%		
NUE				36%																	14%	64%						
NUE_Prot		18%							73%	18%	18%	9%									45%		18%					
NupEFlo								14%					71%			14%				14%				57%				14%
NupEMat							30%	50%		20%		10%		10%		60%			10%				20%	10%				
NutE									67%	17%							33%	83%				33%			33%			
NutE_Prot				31%						13%							56%				13%							
PH	7%		29%					7%	14%		21%							14%		14%								
REMN	25%				42%			17%					25%	17%		8%			33%	8%								17%
SA						9%															18%							
TKW												11%				22%											11%	
%N_FLO			13%	25%			50%				25%	63%	25%													13%		
%N_S		5%	19%			14%	19%	5%					5%	10%					5%					10%				

Supplementary data 6: Empirical distribution of betweenness centrality based on 500 randomizations of the complete colocalisation network. The distribution fits a gamma distribution (shape= 2.169, rate= 0.079). Then this distribution was used to test betweenness centrality. P-value for INN_FLO, FLO, NutE, and %N_Flo are respectively: 0.005, 0.028, 0.035, and 0.039.

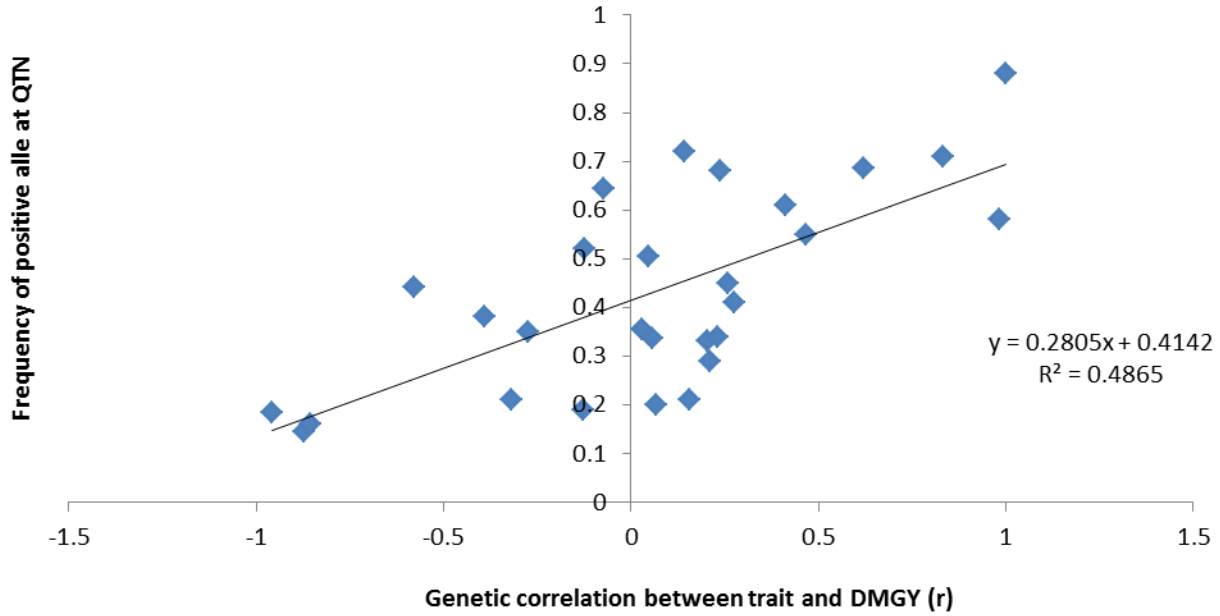


Supplementary data 7: (A) Boxplot of allele frequencies of the alleles which had a positive effect on traits. (B) Median frequency of positive effect allele at QTN as a function of the correlation (r) between traits and yield (DMGY) genetic values. Only varieties registered after 2005 were used (100 varieties).

(A)

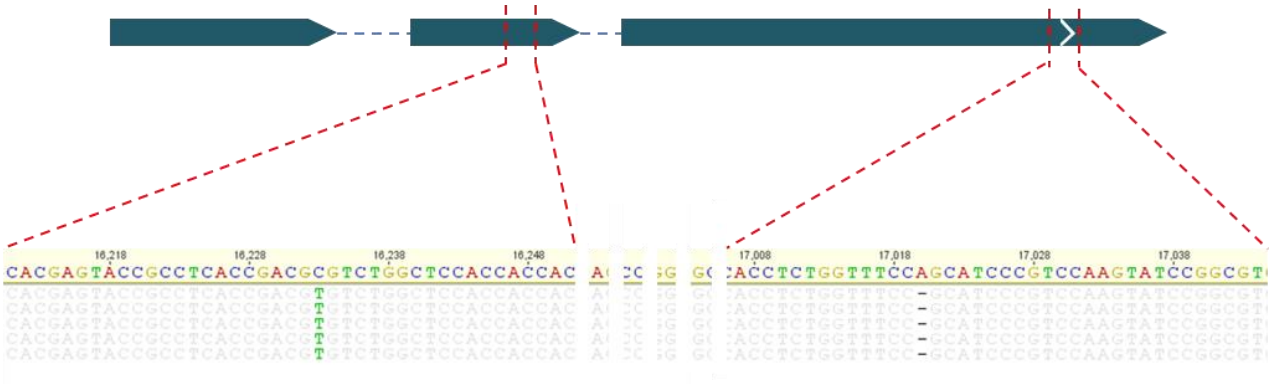


(B)



Supplementary data 8: SNP detection in *NAM-A1*. (A) Gene model Traes_6AS_6F89CC969.1 generated by MIPS (http://pgsb.helmholtz-muenchen.de/plant/wheat/iwgsc/index.jsp) and visualisation of SNP. (B) SNPs context sequences. In SNP2 the deletion has been transformed in A/G to facilitate scoring. **(C) Linkage disequilibrium between SNPs on *NAM-A1* and iSelect 90K SNPs.** Position refers to Wang et al. 2014 genetic map.

(A)



(B)

SNP1 (6AS:4397602_16233)

GAGAAGCTCGGCGTCAAGAAGGCGCTCGTCTTCTACCGCGGGAAGCCGCCCAAGGGCCTCAAAACCAA
CTGGATCATGCACGAGTACCGCCTCACCGACG**[C/T]**GTCTGGCTCCACCACCACCAGCCGGCCGCCGCCG
CCTGTGACCGGCGGGAGCCGGGCTGCAGCCTCTCTGAGGGTACGTACACGTGTCGATCGCACGGTA

SNP2 (6AS:4397602_17020)

CATTTATGAATCCTCTCCCCGTGCAAGACGGGACGTACCATCAACACCATGTCATCCTCGGCGCCCCACT
GGCGCCAGAGGCTACCACAGGCGGCGCCACCTCTGGTTTCC**[A/G]**CATCCCGTCCAAGTATCCGGCGTG
AACTGGAATCCCTGAGCAAATGATATGAACACCACATACGCGCATGCACGCATGCATAACTTTTGCAAG
TGTAGCCAGTAGTTGTTGCAGTTCGTGGTAGTCGCTTTCAG

(C)

NAM-A1	SNP (90K)	LD(r^2)	Chr	Position
	Ra_c28284_223	0.963	6A	74.24
	Tdurum_contig51717_1463	0.963	6A	74.24
SNP1	Tdurum_contig51717_1582	0.963	6A	74.24
	BS00010811_51	0.927	6A	74.24
	BS00010441_51	0.819	6A	74.24
	Kukri_c9595_242	0.781	6A	74.24
	wsnp_Ex_rep_c67878_66584488	0.768	6A	74.24
SNP2	BS00084846_51	0.764	6A	74.24
	wsnp_Ex_c35465_43610634	0.764	6A	74.24
	Kukri_c22893_1651	0.755	6A	74.24

Supplementary data 9: Prediction of NAM-A1 protein sequence: (A) *NAM-A1* coding DNA sequence (CDS), (B) NAM-A1 protein sequence. Prediction made using FGENESH 2.6 (Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters, Genome Biol. 2006,7, Suppl. 1: P. 10.1-10.12). Highlighted, use in 3D conformation; Underlined, NAC domain; **black and bold**, putative DNA binding site; **red**, variation.

(A)

ATGAGGTCCATGGGCAGCTCCGACTCATCCTCCGGCTCGGCGCAAAAAGCAGCGCGGCAT
CAGCATGAGCCGCCGCCTCCGCGGCAGCGGGGCTCGGCGCCGGAGCTCCCACCGGGCTTC
CGGTTCCACCCGACGGACGAGGAGCTGGTCGTGCACTACCTCAAGAAGAAGGCCGCCAAG
GTGCCGCTCCCCGTCACCATCATCGCCGAGGTGGATCTCTACAAGTTCGACCCATGGGAG
CTCCCCGAGAAGGCGACCTTCGGGGAGCAGGAGTGGTACTTCTTCAGCCCGCGCGACCGC
AAGTACCCCAACGGCGCG**CGG**CCGAAC**CGG**GCGGCGACGTCGGGCTACTGGAAGGCCACC
GGCACGGACAAACCTATCCTGGCCTCGGGGACGGGGTGCGGCCTGGTCCGGGAGAAGCTC
GGCGTCAAGAAGGCGCTCGTCTTCTACCGCGGGAAGCCGCCCAAGGGCCTCAAAACCAAC
TGGATCATGCACGAGTACCGCCTCACCGACG**[A/C]**GTCTGGCTCCACCACCACCAGCCGGCCG
CCGCCGCCTGTGACCGGCGGGAGCCGGGCTGCAGCCTCTCTGAGGTTGGACGACTGGGTG
CTGTGCCGCATCTACAAGAAGATCAACAAGGCCGCGGCCGGAGATCAGCAGAGGAGCACG
GAGTGCGAGGACTCCGTGGAGGACGCGGTCACCGCGTACCCGCTCTATGCCACGGCGGGC
ATGGCCGGTGCAGGTGCGCATGGCAGCAACTACGCTTCACCTTCACTGCTCCATCATCAG
GACAGCCATTTCCTGGAGGGCCTGTTCACAGCAGACGACGCCGGCCTCTCGGCGGGCGCC
ACCTCGCTGAGCCACCTGGCCGCGGCGGCGAGGGCGAGCCCGGCTCCGACCAAACAGTTT
CTCGCCCCGTCGTCTTCAACCCCGTTCAACTGGCTCGATGCGTCACCCGCCGGCATCCTG
CCACAGGCAAGGAATTTCCCTGGGTTTAACAGGAGCAGAAACGTCGGCAATATGTCGCTG
TCATCGACGGCCGACATGGCTGGCGCGGCCGGCAATGCGGTGAACGCCATGTCCGCATTT
ATGAATCCTCTCCCCGTGCAAGACGGGACGTACCATCAACACCATGTCATCCTCGGCGCC
CCACTGGCGCCAGAGGCTACCACAGGCGGCGCCACCTCTGGTTTCC**[A/-]**GCATCCCGTCCAA
GTATCCGGCGTGAACTGGAATCCCTGA

(B)

MRSMGSSDSSSGSAQKAARHQHEPPPPRQRGSAPELPPGFRFHPTDEELVVHYLKKKAAK
VPLPVTIIAEVDLYKFDPWELPEKATFGEQEWYFFSPRDRKYPNGA**R**PN**R**AATSGYWKAT
GTDKPILASGTGCGLVREKLGVKKALVFYRGKPPKGLKTNWIMHEYRLTD**[A/V]**SGSTTTSRP
PPPVTGGSRAAASLRLDDWVLCRIYKKINKAAAGDQQRSTECEDSVEDAVTAYPLYATAG
MAGAGAHGSNYASPSLLHHQDSHFLEGLFTADDAGLSAGATSLSHLAAAARASPAPTKQF
LAPSSSTPFNWLDASPAGILPQARNFPGFNRSRNVGNMSLSSTADMAGAAGNAVNAMSAF
MNPLPVQDGTYHQHHVILGAPLAPEATTGGATSGF**[QHPVQVSGVNWNP or RIPSKYPA]**

Supplementary data 10: Khi² test for the observed haplotypes frequencies from the two SNP frequencies for both collections together. Frequencies of each SNP in two collections of bread wheat genotypes (CC = 367-core collection, elite = 334-elite collection), observed and theoretical number of lines for each haplotype in both collection and Khi² test.

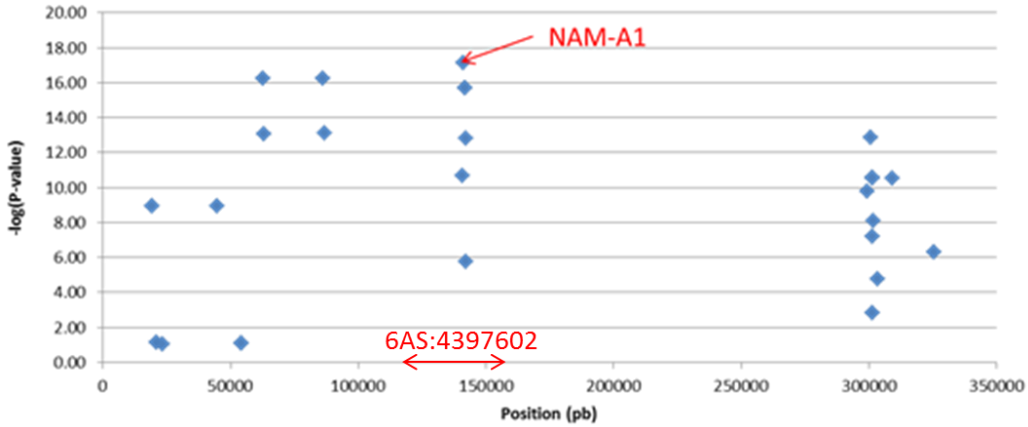
		Frequency		
		CC	**Elite**	**Total**
SNP1	C	0.253	0.085	**0.170**
	T	0.747	0.915	**0.830**
SNP2	A	0.765	0.276	**0.524**
	Del	0.235	0.724	**0.476**

Observed	C	T
A	113	238
Del	1	318

Theoretical	C	T
A	60	291
Del	54	265

Khi²	C	T
A	47.5	9.7
Del	52.3	10.7
	Total	120.3
	Proba	5.4604E-28

Supplementary data 11: Evolution of SNP significance in *NAM-A1* chromosomal region. Phenotyping values of Cormier et al. (2014) were used. 196 elite European varieties were used and SNP effects were tested using the following naïve model: NUE = μ + E + SNP + e. *NAM-A1* chromosomal region was rebuilt by M. Throude.



Supplementary data 12: Protein sequence alignment using ClustalW.

SNP1

```
NAM-A1b                          YRGKPPKGLKTNWIMHEYRLTDASG--STTTSRPPPP--VTGGSRAAASL
NAM-A1d                          YRGKPPKGLKTNWIMHEYRLTDVSG--STTTSRPPPP--VTGGSRAAASL
NAM-A1c                          YRGKPPKGLKTNWIMHEYRLTDVSG--STTTSRPPPP--VTGGSRAAASL
NAM-A1a                          YRGKPPKGLKTNWIMHEYRLTDASG--STTTSRPPPP--VTGGSRAAASL
TaNAM-D1_AIZ97667.1              YRGKPPKGLKTNWIMHEYRLTDASG--STTTSRPPPPPPVTGGSRAAASL
AtNAM-D1_ABI94354.1              YRGKPPKGLKTNWIMHEYRLTDASG--STTTSRPPPPPPVTGGSRAAASL
ttNAM-B1_A0SPJ4.1                YRGKPPKGLKTNWIMHEYRLTDASG--STTATNRPPP--VTGGSRAAASL
TiNAM-B1_AGH32788.1              YRGKPPKGLKTNWIMHEYRLTDASG--STTATNRPPP--VTGGSRAAASL
ttNAM-A2_AIW49540.1              YRGKPPKGLKTNWIMHEYRLTDASS--SATTSRPPPVT----GGSRAASL
TaNAM-D2_AIZ97668.1              YRGKPPKGLKTNWIMHEYRLTDASS--SATTSRPPPVT----GVSRAASL
ttNAM-B2_A0SPJ6.1                YRGKPPKGLKTNWIMHEYRLTDASS--SATTSRPPPVT----GGSRSASL
HvNAM-2_A0SPJ9.1                 YRGKPPKGLKTNWIMHEYRLTDASS--SAATSRPPPVT----GGSRAASL
HvNAM-B1_ACL31422.1              YRGKPPRGLKTNWIMHEYRLTGASA--GSTTTSRPPP--VTGGSRAPASL
HvNAM-1_A0SPJ8.1                 YRGKPPRGLKTNWIMHEYRLTGASA--GSTTTSRPPP--VTGGSRAPASL
Os07g37920_ONAC010_Q8H4S4.1      YRGKPPKGVKTNWIMHEYRLTDTSSSAAAVATTRRPPPPITGGSKGAVSL
AtNAM_ANAC018_Q9ZNU2.1           YSGKPPKGVKSDWIMHEYRLTD-NKP---THICDFGNK--------KNSL
ATNAC2_AEE75684.1                YSGKPPKGVKSDWIMHEYRLIE-NKPNNRPPGCDFGNK--------KNSL
AtNAC025_Q8GY42.1                YGGKPPKGIKTDWIMHEYRLTDGNLSTAAKPPDLTTTR--------KNSL
SNAC1_AIX03023.1                 YAGKAPRGVKTDWIMHEYRLADAGRAAAGAK---------------KGSL
ANAC_ANAC018_Q9C932.1            YIGKAPKGTKTNWIMHEYRLIEPSR--------------------RNGST
```

SNP2

```
NAM-A1b                          NPLPVQDGTYHQHHVILGAPLAPEATTGGATSGFRIPSKYPA------
NAM-A1d                          NPLPVQDGTYHQHHVILGAPLAPEATTGGATSGFRIPSKYPA------
NAM-A1c                          NPLPVQDGTYHQHHVILGAPLAPEATTGGATSGFQHPVQVSGVNWNP-
NAM-A1a                          NPLPVQDGTYHQHHVILGAPLAPEATTGGATSGFQHPVQVSGVNWNP-
TaNAM-D1_AIZ97667.1              NPLPVQDGTYHQHHVILGAPLAPEATAGAATSGFQHHAVQISGVNWNP
AtNAM-D1_ABI94354.1              NPLPVQDGTYHQHHVILGAPLAPEATAGAATSGFQHHAVQISGVNWNP
ttNAM-B1_A0SPJ4.1                TYLPVQDGTYHQQHVILGAPLVPEAAA--ATSGFQHPVQISGVNWNP-
TiNAM-B1_AGH32788.1              TYLRVQDGTYHQQHVILGAPLVPEAAA--ATSGFQHPVQISGVNWNP-
ttNAM-A2_AIW49540.1              NHLPVQDGTYHQQHVILGTPLAPEATA-AATSAFQHPVQISGVNWNP-
TaNAM-D2_AIZ97668.1              SHLPVQDGTYHQQHVILGAPLAPEATA-AATSAFQHPVQISGVNWNP-
ttNAM-B2_A0SPJ6.1                NHLPMQDGTYHQQHVILGAPLAPEATA-AATSAFQHPVQISGVNWNP-
HvNAM-2_A0SPJ9.1                 NHLPVQDGTYHQQHVILGAPLAPEATG-AAASAFQHPVQISGVNWNP-
HvNAM-B1_ACL31422.1              MYLPVQDGTYHQHVILG-APLAPEAIAGAATSGFQHHVQISGVNWNP-
HvNAM-1_A0SPJ8.1                 MYLPVQDGTYHQHVILG-APLAPEAIAGAATSGFQHHVQISGVNWNP-
Os07g37920_ONAC010_Q8H4S4.1      NPLGVQGATYQQHQAIMGASLPSESAAAAAACNFQHPFQLSRVNWDS-
AtNAM_ANAC018_Q9ZNU2.1           ---------DCSTSMAATPLMQNQG----------GIYQLPGLNWYS-
ATNAC2_AEE75684.1                ---GDCSNMSSSMMEETPPLMQQQGGVLGDGLFRTTSYQLPGLNWYSS
AtNAC025_Q8GY42.1                ---PQSSGFHANGVMDTTSSLADHG-------VLRQAFQLPNMNWHS-
SNAC1_AIX03023.1                 --------MYSGLDMLPPGDDFYSSLFASPRVKGTTPRAGAGMGMVPF
ANAC_ANAC018_Q9C932.1            -YLKTEEEVESSHGFNNSGELAQKGYG---VDSFGYSGQVGGFGFM--
```

SUPPORTING INFORMATION ON PART IV

[Supplementary data of Cormier et al. (2015) Improving genomic prediction using a GWAS-based method to pre-select marker in multi-environment data. Expected submission: April 2015 **and** on epistatic interactions]

Figure S1: Evolution of accuracy variance for predicting the genetic values of NUE and NHI. Predictions were assessed using a three-fold design repeated 50 times for each combination of SNP number and SNP section rank in MET-GWAS-based ranking. Red arrows highlight the optimum.

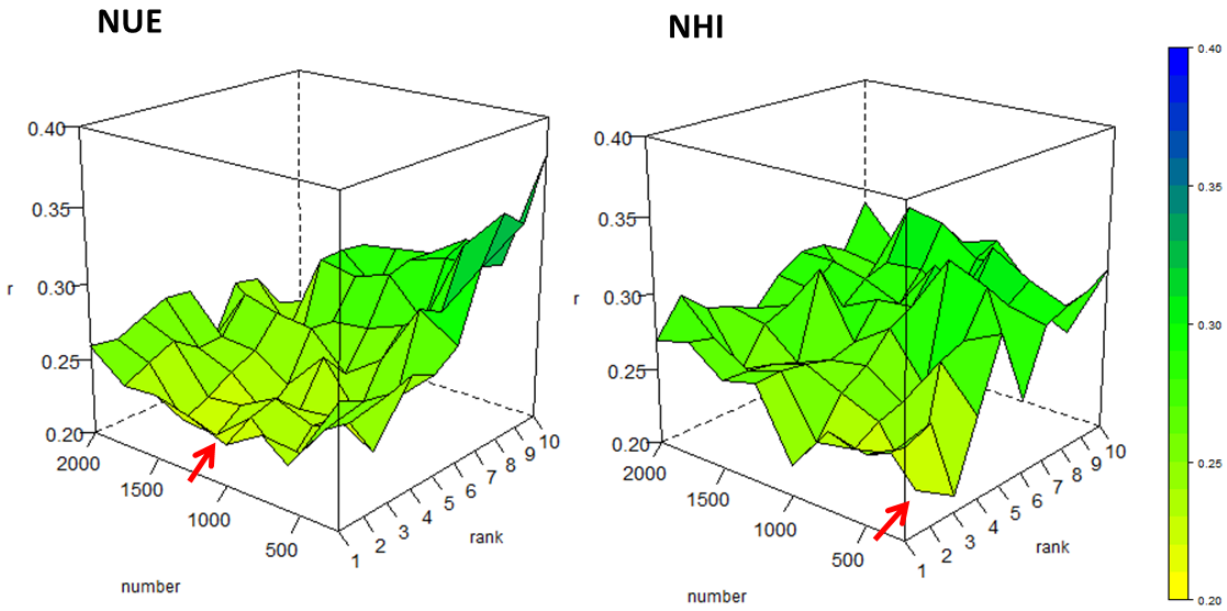


Figure S2: Significance of the difference in accuracy between the optimum and the other combination of section size and section rank. NUE optimum: 1250 SNPs, section rank 1; NHI optimum: 500 SNPs section rank 3.A Wilcoxon test was performed. Plotted values are -$\log_{10}$(P-value).

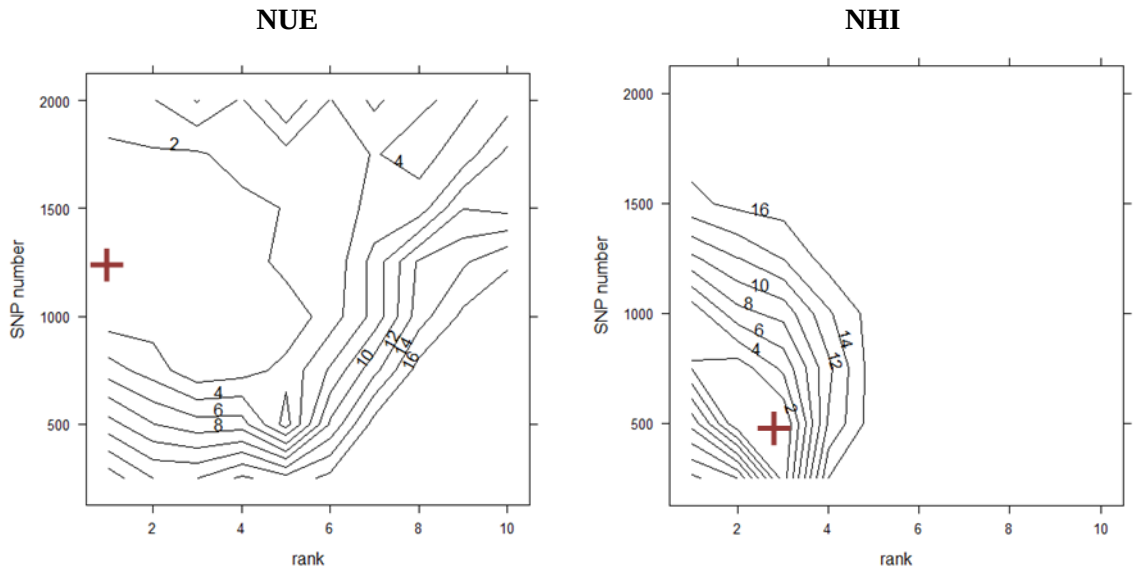


Figure S3: Significance of the difference in accuracy between the optimum and the other combination of section size and section rank for G×E interaction prediction in (1) CV1, (2) CV2, and (3) CV3 for (A) NUE and (B) NHI. In CV1 and CV2, predictions were assessed using a three-fold design repeated 50 times. In CV3, a four-fold design was repeated 28 times. NUE optimum: 500 SNPs, section rank 1; NHI optimum: 250 SNPs, section rank 1.A Wilcoxon test was performed. Plotted values are -$\log_{10}$(P-value).

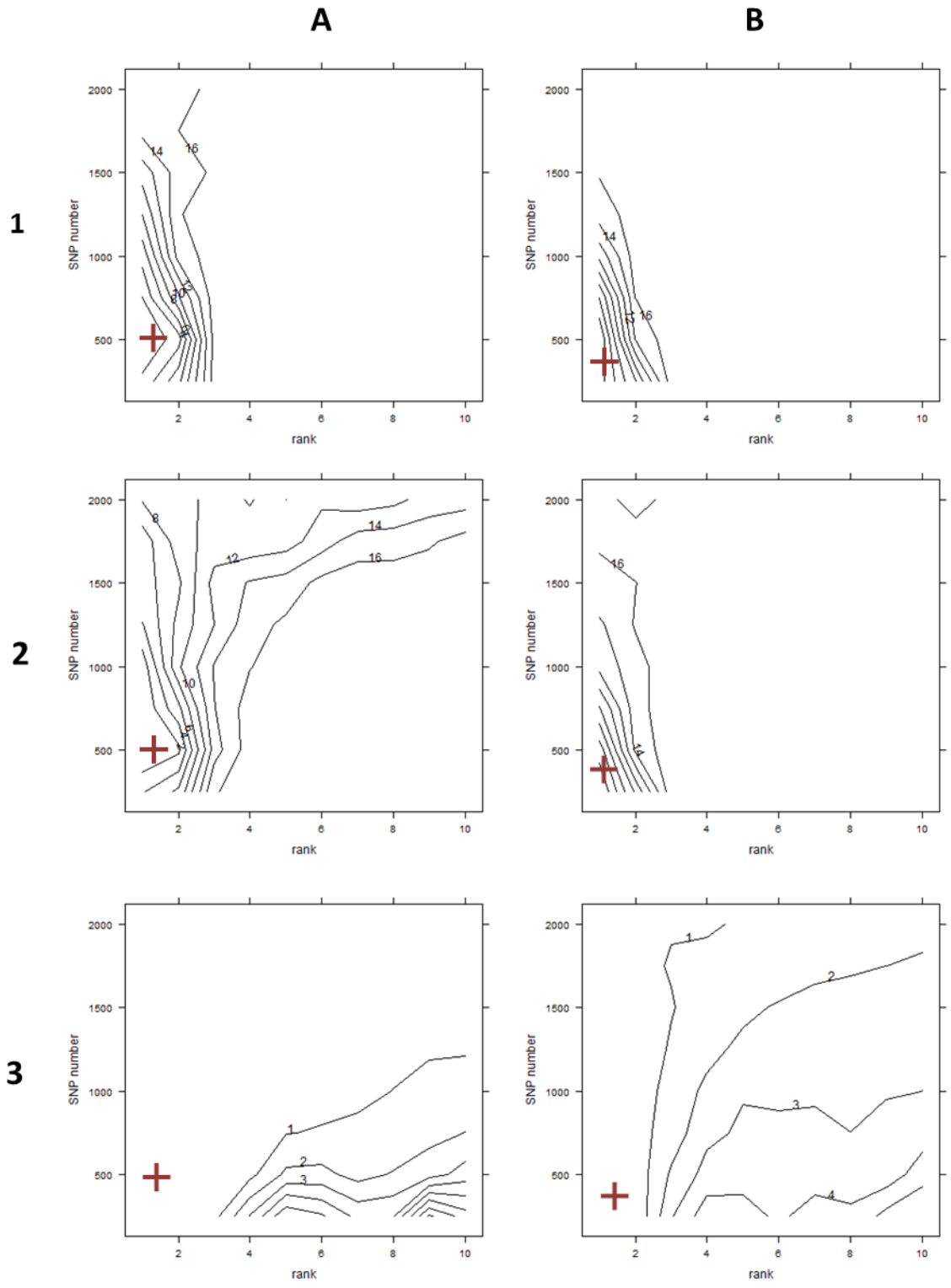


Figure S4: SNP rank in function of -$\log_{10}$(P). Graphs represent values for both NUE and NHI, for the SNP additive effect (α_j) and for the most significant interaction between SNPs and ECs (β_j). Red points represent SNPs that were used at optimum. For β_j, -$\log_{10}$(P) start at -$\log_{10}$(0.05) =1.3 as this threshold was used to stop adding ECs in the MET-GWAS model (forward approach).

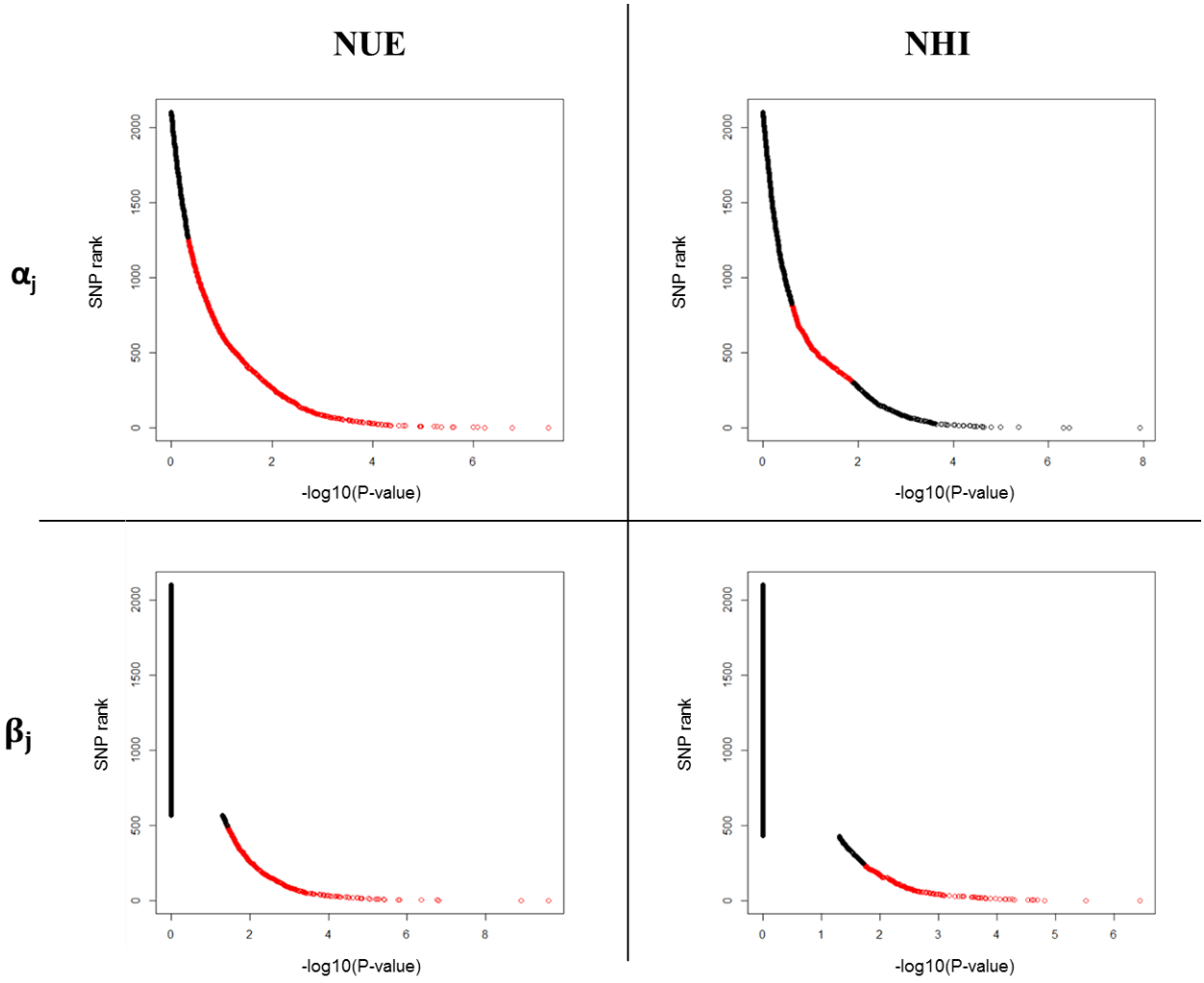


Figure S5: Correlation of G × E interactions between environments for NUE and NHI. Values are pairwise correlations (r).

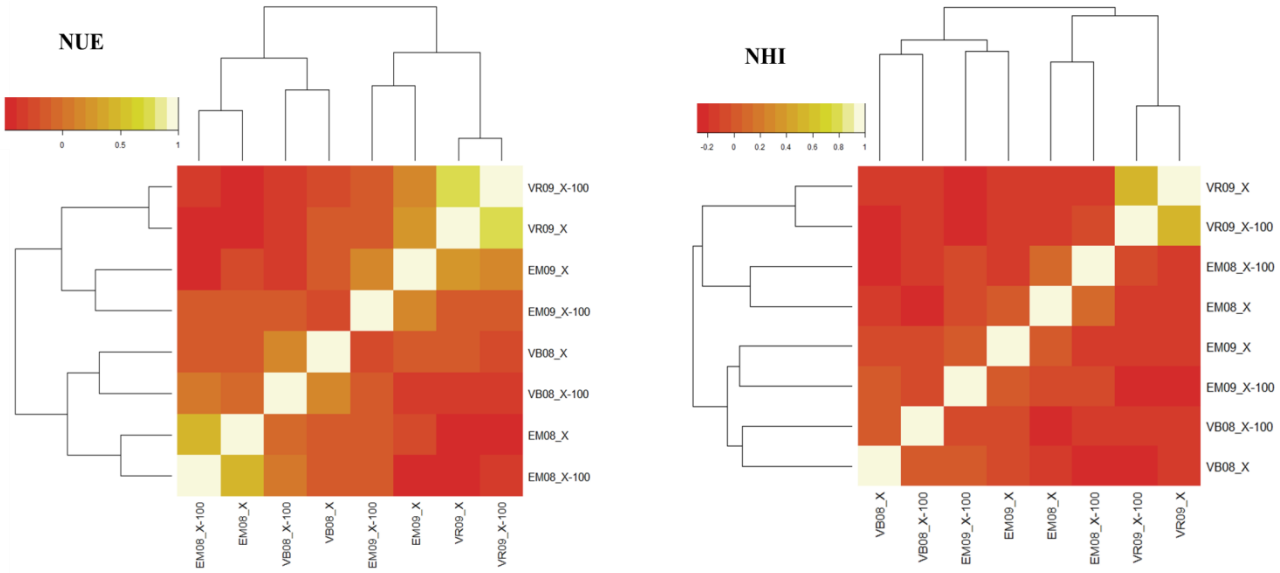


Figure S6: Heatmap of Ω the environment covariance matrix based on ECs used to estimated G×E interactions in genomic prediction models.

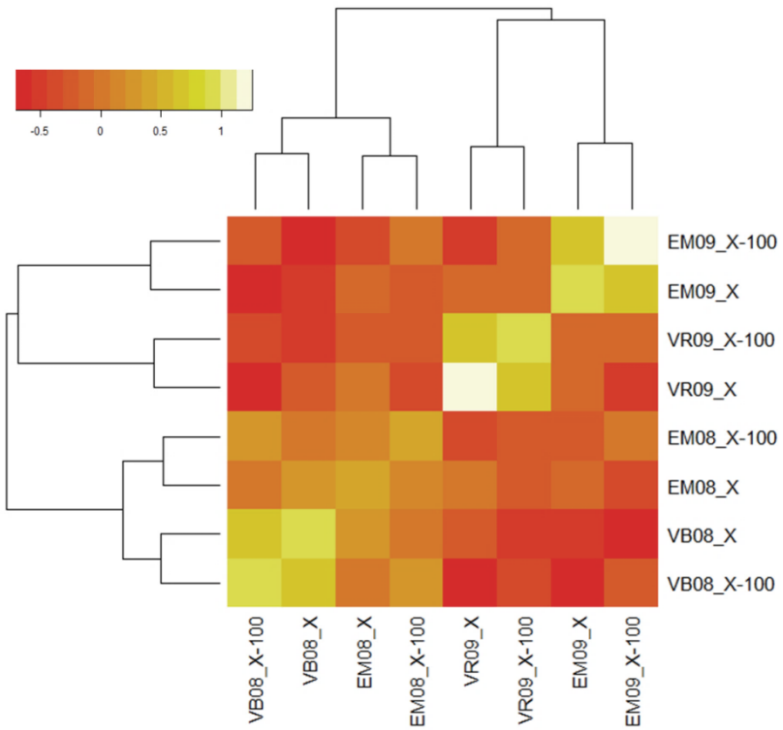


Table S1: Comparison of accuracies adding G×E predictions and pre-selecting SNPs. The number of SNPs used to compute in matrices $\mathbf{K_1}$ and $\mathbf{K_2}$ (models (6) and (7); See Materials and Methods) are indicated in columns $\mathbf{K_1}$ and $\mathbf{K_2}$. When all available SNPs or all SNP that were pre-selected based on LD were used, $\mathbf{K_1}$=$\mathbf{K_2}$. $r_{(yijk\text{-}ej/gi)}$ and $r_{(yijk\text{-}ej/gi+gwij)}$ are prediction accuracies of models (6) and (7), respectively.

	CV	Optimum				Pre-select on LD			Opt/LD[2]	All SNPs			Opt/All[3]
		$\mathbf{K_1}$	$\mathbf{K_2}$	$\mathbf{r_{(yijk\text{-}Ej/gi)}}$	$\mathbf{r_{(yijk\text{-}Ej/gi+gwij)}}$[a]	$\mathbf{K_1 = K_2}$	$\mathbf{r_{(yijk\text{-}Ej/gi)}}$	$\mathbf{r_{(yijk\text{-}Ej/gi+gwij)}}$[b]		$\mathbf{K_1 = K_2}$	$\mathbf{r_{(yijk\text{-}Ej/gi)}}$	$\mathbf{r_{(yijk\text{-}Ej/gi+gwij)}}$[c]	
NUE	1	1,250	500	0.53+/-0.07	0.61+/-0.05***	2,101	0.50+/-0.07	0.53+/-0.06**	***	25,368	0.48+/-0.06	0.52+/-0.06**	***
	2	250		0.63+/-0.02	0.72+/-0.02***		0.65+/-0.02	0.72+/-0.02***	ns.		0.64+/-0.02	0.71+/-0.02***	ns.
	3			0.61+/-0.07	0.66+/-0.14*		0.63+/-0.07	0.68+/-0.10*	ns.		0.63+/-0.07	0.67+/-0.10*	ns.
NHI	1	500	250	0.34+/-0.04	0.44+/-0.04***		0.19+/-0.06	0.24+/-0.05***	***		0.22+/-0.05	0.25+/-0.05***	***
	2	250		0.35+/-0.02	0.46+/-0.03***		0.34+/-0.02	0.41+/-0.03***	***		0.35+/-0.02	0.41+/-0.03***	***
	3			0.31+/-0.06	0.36+/-0.12*		0.32+/-0.06	0.34+/-0.12 ns.	ns.		0.32+/-0.06	0.34+/-0.12 ns.	ns.

[a]Result of the Wilcoxon test between $r_{(yijk\text{-}Ej/gi)}$ and $r_{(yijk\text{-}Ej\ /gi+gwij)}$
[b]Result of the Wilcoxon test between the optimum and the use of all SNPs pre-selected based on LD for the complete model
[c]Result of the Wilcoxon test between the optimum and the use of all available SNPs for the complete model

***: P-value <0.001 ; **: P-value <0.01; *: P-value <0.05; and ns.: non-significant P-value>0.05

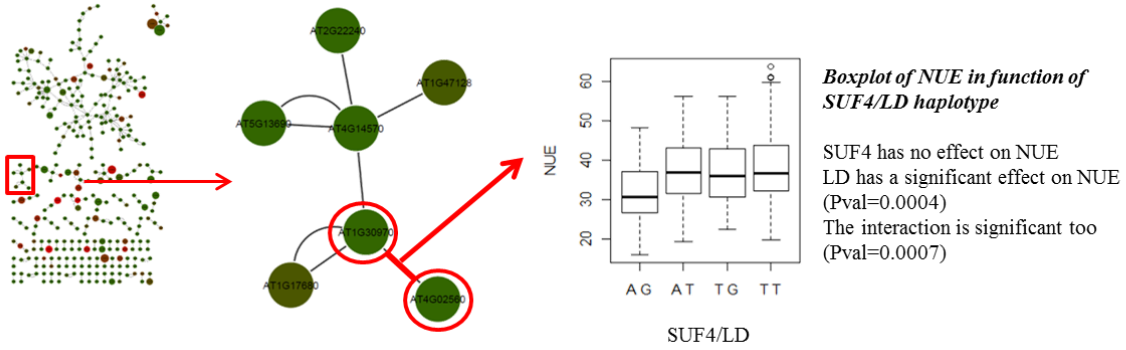
Table S2: Description of environmental covariates (ECs) used to predict G×E interactions.

Stress	EC	Description
Nitrogen	NTA_Max	Estimation of the N available as in Cormier et al. (2013)
	NSupply	Total N supply
	NResidual	Residual soil N
	N_End_Wint	N supply at the end of winter
	N_Z30	N supply at Z30
	N_Z32	N supply at Z32
Frost	Nbrj_Tmin<-4	Number of days with a minimal temperature < - 4°C
	Sum_Tmin	Sum of daily temperature< - 4°C
Radiation	Deficit_Rg	Number of days with global radiation < 1045J/cm² during meiosis +/-5 days
	Sum_deficit_Rg	Sum of global radiation < 1045J/cm² during the all crop cycle
	Sum_Rg	Sum of global radiation during the all crop cycle
Heat	Stress_Tmax>25	Number of days with a maximal temperature > 25°C
	Sum_Tmax>25	Sum of daily temperature > 25°C
Drought	Sum_Rain	Sum of daily rainfall during the all crop cycle
	Nbrj_P<ETP	Number of days with a potential evapotranspiration > rainfall during the all crop cycle
	Moy_NbrjP<ETP	Mean of the number of consecutive day with a potential evapotranspiration > rainfall during the all crop cycle
	Mean_DeficitH2O	Mean of the cumulative hydric deficit (daily sum of potential evapotranspiration - rainfall) during the all crop cycle
	Thrmq_R	Mean of the sum of temperature >0°C divided by sum of global radiation during the all crop cycle

Table S3: ECs values by environment.

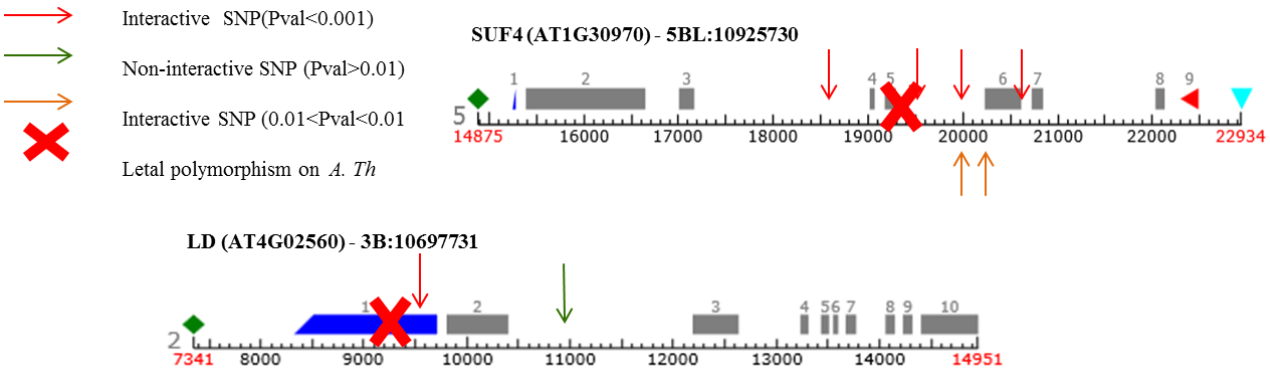
	Site_Season	VB08		VR09		EM08		EM09	
	N levels	**X-100**	**X**	**X-100**	**X**	**X-100**	**X**	**X-100**	**X**
Nitrogen	NTA_Max	157.31	242.12	172.73	236.49	143.62	205.97	110.9	241.34
	NSupply	150.0	232.5	150.0	250.0	137.0	237.0	80.0	180.0
	NResidual	106	106	30	30	67	67	30	30
	N_End_Wint	0	0	60	60	0	50	0	50
	N_Z30	44.0	66.5	60.0	100.0	70.0	70.0	50.0	50.0
	N_Z32	0	60	0	60	0	50	0	50
Frost	Nbrj_Tmin<-4	7	7	17	17	7	7	14	14
	Sum_Tmin	-5.2	-5.2	-50.5	-50.5	-4.6	-4.6	-59.7	-59.7
Radiation	Deficit_Rg	1	1	1	1	1	1	2	2
	Sum_deficit_Rg	89 636	89 636	90 254	90 254	89 943	89 943	96 572	96 572
	Sum_Rg	300 508	300 508	339 037	339 037	320 091	320 091	320 696	320 696
Heat	Stress_Tmax>25	15	15	31	31	20	20	23	23
	Sum_Tmax>25	24	24	102	102	39	39	58	58
Drought	Sum_Rain	487	487	525	525	493	493	390	390
	Nbrj_P<ETP	150	150	152	152	163	163	173	173
	Moy_NbrjP<ETP	2	2	2	2	3	3	3	3
	Mean_DeficitH2O	-140	-140	-109	-109	-97	-97	-54	-54
	Thrmq_R	88	88	105	105	94	94	96	96

Supplementary data 10: Example of "validated" epistatic interaction. The case of SUF4 and LD



Kim et al. (2006) SUPPRESSOR OF FRIGIDA4, Encoding a C2H2-Type Zinc Finger Protein, Represses Flowering by Transcriptional Activation of Arabidopsis FLOWERING LOCUS C. Plant Cell 18:1985-2998

SUF4 binds to FRI and FRL1 to form a complex which is the transcriptional activator of FLC.
When FRI is absent, LD bind to SUF4 (two yeast assay + validation in plant) and suppresses SUF4 activity



NITROGEN USE EFFICIENCY IN BREAD WHEAT (T. AESTIVUM L.): BREEDING & GENE DISCOVERY - **Fabien Cormier - 27 mai 2015**

SUMMARY: In a context of fertiliser reduction, breeding for enhanced nitrogen use efficiency in bread wheat is necessary. This PhD thesis resulting from private-public collaboration between the French National Institute for Agricultural Research and Biogemma aimed providing necessary tools. Analyses were conducted using a dataset of 225 commercial varieties genotyped with 24K SNP and tested in eight combinations of year, location, and nitrogen regimes. We showed that even if past selection increased nitrogen use efficiency at high and moderate nitrogen regimes, genetic progresses need to be accelerated and better balanced between traits. This could be achieved by mixing phenotypic and marker assisted selections. In this sense, we developed a method to define quantitative trait locus from genome-wide association study: 333 chromosomal regions involved in 28 NUE-related traits have been identified. The *NAM-A1* gene was located in one of these regions and its natural variants were characterized. We also showed that genomic selection could be improved by pre-selecting SNP based on their significance in a multi-environmental genome-wide association study. Networks of epistasis interactions were also studied and an interesting sub-network was identified. Results and methods are discussed regarding breeding and gene discovery strategy. Further investigations and improvements are suggested.

Keyword: Epistasis, GWAS, Genomic selection, *NAM-A1*, Nitrogen, Quantitative genetics, *Triticum aestivum* (L.), Wheat

---

RESUME: Dans un contexte de réduction des intrants agricoles, la création de variétés de blé qui utilisent l'azote de manière plus efficiente est aujourd'hui nécessaire. Cette thèse, issue d'un partenariat public-privée entre l'institut nationale de la recherche agronomique et Biogemma, avait pour but d'apporter des outils nécessaires à la création de variétés répondant à cette exigence. Pour ce faire, nous avons analysé 225 variétés commerciales génotypées avec 24K SNP et testées dans huit combinaisons d'année, lieu et régime azoté. Nous avons montré que même si la sélection a amélioré l'efficience d'utilisation de l'azote en condition optimale et sub-optimale, ce progrès génétique doit être accéléré et mieux réparti entre les différents traits. Nous proposons pour cela de mixer sélection phénotypique et sélection assistée par marqueurs. Dans ce sens, nous avons développé une méthode pour définir les régions chromosomiques associées à nos 28 traits. Parmi les 333 régions identifiées, nous avons notamment localisé le gène *NAM-A1* et avons pu caractériser ses variants naturels. Nous avons aussi montré que la sélection génomique pourrait être plus efficace si les SNP étaient présélectionnés en fonction de leurs significativités en génétique d'association multi-environnementale. Les réseaux d'interactions épistatiques furent aussi étudiés, mettant en évidence un sous-réseau particulièrement intéressant. Nos résultats et méthodes sont discutés au regard des stratégies d'amélioration variétale et de découverte de gènes. Des pistes de recherche complémentaires et des améliorations ont aussi été suggérées.

Mots-clés: Azote, Blé, Epistasie, Génétiques quantitative, GWAS, *NAM-A1*, Sélection génomique, *Triticum aestivum* (L.)